



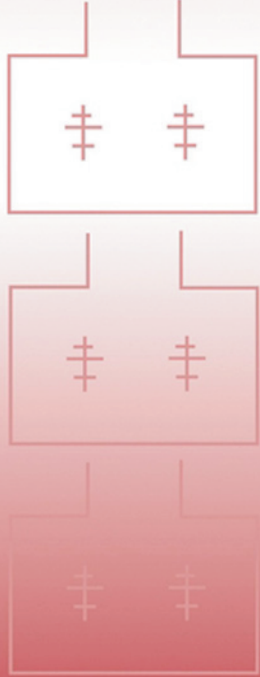
 WILEY

Design and Analysis of Experiments

Volume 1
Introduction to Experimental Design

SECOND EDITION

Klaus Hinkelmann
Oscar Kempthorne



Design and Analysis of Experiments



THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

Design and Analysis of Experiments

Volume 1

Introduction to Experimental Design

Second Edition

Klaus Hinkelmann

Virginia Polytechnic Institute and State University
Department of Statistics
Blacksburg, VA

Oscar Kempthorne

Iowa State University
Department of Statistics
Ames, IA



WILEY-INTERSCIENCE
A John Wiley & Sons, Inc., Publication

Copyright © 2008 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Wiley Bicentennial Logo: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data:

Hinkelmann, Klaus, 1932–

Design and analysis of experiments / Klaus Hinkelmann, Oscar Kempthorne. — 2nd ed.

v. cm. — (Wiley series in probability and statistics)

Includes index.

Contents: v. 1. Introduction to experimental design

ISBN 978-0-471-72756-9 (cloth)

1. Experimental design. I. Kempthorne, Oscar. II. Title.

QA279.K45 2008

519.5'7—dc22

2007017347

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Preface to the Second Edition	xvii
Preface to the First Edition	xxi
1 The Processes of Science	1
1.1 INTRODUCTION	1
1.1.1 Observations in Science	1
1.1.2 Two Types of Observations	2
1.1.3 From Observation to Law	3
1.2 DEVELOPMENT OF THEORY	5
1.2.1 The Basic Syllogism	5
1.2.2 Induction, Deduction, and Hypothesis	6
1.3 THE NATURE AND ROLE OF THEORY IN SCIENCE	8
1.3.1 What Is Science?	8
1.3.2 Two Types of Science	9
1.4 VARIETIES OF THEORY	11
1.4.1 Two Types of Theory	11
1.4.2 What Is a Theory?	12
1.5 THE PROBLEM OF GENERAL SCIENCE	14
1.5.1 Two Problems	15
1.5.2 The Role of Data Analysis	15
1.5.3 The Problem of Inference	16
1.6 CAUSALITY	16
1.6.1 Defining Cause, Causation, and Causality	17
1.6.2 The Role of Comparative Experiments	19
1.7 THE UPSHOT	21
1.8 WHAT IS AN EXPERIMENT?	21
1.8.1 Absolute and Comparative Experiments	22
1.8.2 Three Types of Experiments	23
1.9 STATISTICAL INFERENCE	24
1.9.1 Drawing Inference	24
1.9.2 Notions of Probability	25
1.9.3 Variability and Randomization	26

2	Principles of Experimental Design	29
2.1	CONFIRMATORY AND EXPLORATORY EXPERIMENTS	29
2.2	STEPS OF DESIGNED INVESTIGATIONS	30
2.2.1	Statement of the Problem	31
2.2.2	Subject Matter Model	32
2.2.3	Three Aspects of Design	33
2.2.4	Modeling the Response	35
2.2.5	Choosing the Response	36
2.2.6	Principles of Analysis	36
2.3	THE LINEAR MODEL	37
2.3.1	Three Types of Effects	37
2.3.2	Experimental and Observational Units	38
2.3.3	Outline of a Model	40
2.4	ILLUSTRATING INDIVIDUAL STEPS: STUDY 1	41
2.4.1	The Questions and Hypotheses	41
2.4.2	The Experiment and a Model	41
2.4.3	Analysis	42
2.4.4	Alternative Experimental Setup	44
2.5	THREE PRINCIPLES OF EXPERIMENTAL DESIGN	45
2.6	THE STATISTICAL TRIANGLE: STUDY 2	46
2.6.1	Statement of the Problem	46
2.6.2	Four Experimental Situations	46
2.7	PLANNING THE EXPERIMENT: THINGS TO THINK ABOUT	51
2.8	COOPERATION BETWEEN SCIENTIST AND STATISTICIAN	53
2.9	GENERAL PRINCIPLE OF INFERENCE AND TYPES OF STATISTICAL ANALYSES	56
2.9.1	General Model	56
2.9.2	Outline of the ANOVA	56
2.10	OTHER CONSIDERATIONS FOR EXPERIMENTAL DESIGNS	58
3	Survey of Designs And Analyses	61
3.1	INTRODUCTION	61
3.2	ERROR-CONTROL DESIGNS	62
3.3	TREATMENT DESIGNS	64
3.4	COMBINING IDEAS FROM ERROR-CONTROL AND TREATMENT DESIGNS	65
3.5	SAMPLING DESIGNS	68
3.6	ANALYSIS AND STATISTICAL SOFTWARE	68
3.7	SUMMARY	69
4	Linear Model Theory	71
4.1	INTRODUCTION	71
4.1.1	The Concept of a Model	71
4.1.2	Comparative and Absolute Experiments	73
4.2	REPRESENTATION OF LINEAR MODELS	73
4.3	FUNCTIONAL AND CLASSIFICATORY LINEAR MODELS	74

4.3.1	Functional Models	74
4.3.2	Classificatory Models	74
4.3.3	Models with Classificatory and Functional Components	75
4.4	THE FITTING OF $y = X\beta$	76
4.4.1	The Notion of Identifiability	76
4.4.2	The Notion of Estimability	77
4.4.3	The Method of Least Squares	77
4.4.4	Theory of Linear Equations	81
4.5	MOORE-PENROSE GENERALIZED INVERSE	84
4.6	CONDITIONED LINEAR MODEL	85
4.6.1	Affine Linear Model	85
4.6.2	Normal Equations for the Conditioned Model	87
4.6.3	Different Types of Conditions	88
4.6.4	General Case	89
4.7	TWO-PART LINEAR MODEL	90
4.7.1	Ordered Linear Models	90
4.7.2	Using Orthogonal Projections	91
4.7.3	Orthogonal ANOVA	93
4.8	SPECIAL CASE OF A PARTITIONED MODEL	94
4.9	THREE-PART MODELS	94
4.10	TWO-WAY CLASSIFICATION WITHOUT INTERACTION	95
4.11	K-PART LINEAR MODEL	97
4.11.1	The General Model and Its Sums of Squares	97
4.11.2	The Means Model	99
4.12	BALANCED CLASSIFICATORY STRUCTURES	100
4.12.1	Factors, Levels, and Partitions	101
4.12.2	Nested, Crossed, and Confounded Factors	101
4.12.3	The Notion of Balance	102
4.12.4	Balanced One-Way Classification	102
4.12.5	Two-Way Classification with Equal Numbers	103
4.12.6	Experimental versus Observational Studies	104
4.12.7	General Classificatory Structure	106
4.12.8	The Well-Formulated Model	109
4.13	UNBALANCED DATA STRUCTURES	112
4.13.1	Two-Fold Nested Classification	112
4.13.2	Two-Way Cross-Classification	113
4.13.3	Two-Way Classification without Interaction	116
4.14	ANALYSIS OF COVARIANCE MODEL	118
4.14.1	The Question of Explaining Data	118
4.14.2	Obtaining the ANOVA Table	120
4.14.3	The Case of One Covariate	121
4.14.4	The Case of Several Covariates	121
4.15	FROM DATA ANALYSIS TO STATISTICAL INFERENCE	122
4.16	SIMPLE NORMAL STOCHASTIC LINEAR MODEL	123
4.16.1	The Notion of Estimability	123

4.16.2	Gauss-Markov Linear Model	124
4.16.3	Ordinary Least Squares and Best Linear Unbiased Estimators	126
4.16.4	Expectation of Quadratic Forms	128
4.17	DISTRIBUTION THEORY WITH GMNLM	128
4.17.1	Distributional Properties of $\lambda'\beta$	128
4.17.2	Distribution of Sums of Squares	130
4.17.3	Testing of Hypotheses	131
4.18	MIXED MODELS	132
4.18.1	The Notion of Fixed, Mixed and Random Models	132
4.18.2	Aitken-like Model	133
4.18.3	Mixed Models in Experimental Design	134
5	Randomization	137
5.1	INTRODUCTION	137
5.1.1	Observational versus Intervention Studies	137
5.1.2	Historical Controls versus Repetitions	139
5.2	THE TEA TASTING LADY	139
5.3	TRIANGULAR EXPERIMENT	140
5.3.1	Medical Example	141
5.3.2	Randomization, Probabilities, and Beliefs	141
5.4	SIMPLE ARITHMETICAL EXPERIMENT	142
5.4.1	Noisy Experiments	142
5.4.2	Investigative Experiments and Beliefs	144
5.4.3	Randomized Experiments	145
5.5	RANDOMIZATION IDEAS	148
5.6	EXPERIMENT RANDOMIZATION TEST	150
5.7	INTRODUCTION TO SUBSEQUENT CHAPTERS	151
6	Completely Randomized Design	153
6.1	INTRODUCTION AND DEFINITION	153
6.2	RANDOMIZATION PROCESS	154
6.2.1	Use of Random Numbers	154
6.2.2	Design Random Variables	154
6.3	DERIVED LINEAR MODEL	157
6.3.1	Conceptual Responses and Observations	157
6.3.2	Distributional Properties	159
6.3.3	Additivity in the Broad Sense	161
6.3.4	Error Structure	162
6.3.5	Summary of Results	164
6.4	ANALYSIS OF VARIANCE	165
6.4.1	Deriving the ANOVA Table	165
6.4.2	Obtaining Expected Mean Squares	168
6.5	STATISTICAL TESTS	171
6.5.1	Enumerating Randomizations	171
6.5.2	Randomization Test	172
6.6	APPROXIMATING THE RANDOMIZATION TEST	174

6.6.1	Moments of the Test Statistic	174
6.6.2	Approximation by the F -Test	177
6.6.3	Simulation Study	177
6.7	CRD WITH UNEQUAL NUMBERS OF REPLICATIONS	179
6.7.1	Randomization	180
6.7.2	The Model and ANOVA	180
6.7.3	Comparing Randomization Test and F -Test	180
6.8	NUMBER OF REPLICATIONS	180
6.8.1	Power of the F -Test	182
6.8.2	Smallest Detectable Difference	184
6.8.3	Practical Considerations	185
6.9	SUBSAMPLING IN A CRD	191
6.9.1	Subsampling Model	191
6.9.2	Inferences with Subsampling	193
6.9.3	Comparison of CRDs without and with Subsampling	193
6.10	TRANSFORMATIONS	196
6.10.1	Nonadditivity in the General Sense	196
6.10.2	Nonconstancy of Variances	197
6.10.3	Choice of Transformation	198
6.10.4	Power Transformations	200
6.11	EXAMPLES USING SAS®	201
6.12	EXERCISES	204
7	Comparisons of Treatments	213
7.1	INTRODUCTION	213
7.2	COMPARISONS FOR QUALITATIVE TREATMENTS	213
7.2.1	Treatment Contrasts	214
7.2.2	Orthogonal Contrasts	214
7.2.3	Partitioning the Treatment Sum of Squares	215
7.3	ORTHOGONALITY AND ORTHOGONAL COMPARISONS	218
7.4	COMPARISONS FOR QUANTITATIVE TREATMENTS	219
7.4.1	Comparisons for Treatments with Equidistant Levels	219
7.4.2	Use of Orthogonal Polynomials	220
7.4.3	Contrast Sums of Squares and the ANOVA	223
7.5	MULTIPLE COMPARISON PROCEDURES	224
7.5.1	Multiple Comparisons and Error Rates	224
7.5.2	Least Significant Difference Test	225
7.5.3	Bonferroni t -Statistics	225
7.5.4	Studentized Range Procedure	226
7.5.5	Duncan's Multiple Range Test	226
7.5.6	Scheffé's Procedure	227
7.5.7	Comparisons with a Control	227
7.5.8	Alternatives to Tests Based on Normality	228
7.6	GROUPING TREATMENTS	229
7.7	EXAMPLES USING SAS®	230
7.8	EXERCISES	236

8	Use of Supplementary Information	239
8.1	INTRODUCTION	239
8.2	MOTIVATION OF THE PROCEDURE	240
8.3	ANALYSIS OF COVARIANCE PROCEDURE	242
8.3.1	Basic Model	242
8.3.2	Least Squares Analysis	242
8.3.3	Least Squares Means	244
8.3.4	Formulation in Matrix Notation	245
8.3.5	ANOVA Table	246
8.4	TREATMENT COMPARISONS	250
8.4.1	Preplanned Comparisons	250
8.4.2	Multiple Comparison Procedures	251
8.5	VIOLATION OF ASSUMPTIONS	252
8.5.1	Linear Relationship between x and y	252
8.5.2	Common Slope	253
8.5.3	Covariates Affected by Treatments	256
8.5.4	Normality Assumption	257
8.6	ANALYSIS OF COVARIANCE WITH SUBSAMPLING	258
8.7	CASE OF SEVERAL COVARIATES	259
8.7.1	General Case	260
8.7.2	Two Covariates	262
8.8	EXAMPLES USING SAS®	264
8.9	EXERCISES	274
9	Randomized Block Designs	277
9.1	INTRODUCTION	277
9.2	RANDOMIZED COMPLETE BLOCK DESIGN	278
9.2.1	Definition	278
9.2.2	Derived Linear Model	278
9.2.3	Estimation of Treatment Contrasts	282
9.2.4	Analysis of Variance	282
9.2.5	Randomization Test and F -Test	285
9.2.6	Additivity in the Broad Sense	286
9.2.7	Subsampling in an RCBD	288
9.3	RELATIVE EFFICIENCY OF THE RCBD	288
9.3.1	Question of Effectiveness of Blocking	288
9.3.2	Use of Uniformity Trials	290
9.3.3	Interpretation and Use of Relative Efficiency	291
9.4	ANALYSIS OF COVARIANCE	292
9.4.1	The Model	292
9.4.2	Least Squares Analysis	293
9.4.3	The ANOVA Table	294
9.5	MISSING OBSERVATIONS	295
9.5.1	Estimating a Missing Observation	295
9.5.2	Using the Estimated Missing Observation	297

9.5.3	Several Missing Observations	298
9.6	NONADDITIVITY IN THE RCBD	300
9.6.1	The Problem of Nonadditivity	300
9.6.2	General Model for Nonadditivity	300
9.6.3	One Blocking Factor: A Specific Model for Nonadditivity	302
9.6.4	Testing for Nonadditivity	303
9.6.5	Tukey's Test for Nonadditivity	303
9.6.6	Generalizations	305
9.6.7	Several Blocking Factors	306
9.6.8	Dealing with Block-Treatment Interaction	312
9.7	GENERALIZED RANDOMIZED BLOCK DESIGN	314
9.7.1	Definition	314
9.7.2	Derived Linear Model	314
9.7.3	The ANOVA Table	317
9.7.4	Analyzing Block-Treatment Interaction	319
9.7.5	A More General Formulation	323
9.7.6	Random Block Effects	324
9.7.7	Using Satterthwaite's Procedure	326
9.8	INCOMPLETE BLOCK DESIGNS	328
9.8.1	General Notion of Designs with Incomplete Blocks	328
9.8.2	Balanced Incomplete Block Designs	330
9.8.3	Balanced Treatment Incomplete Block Designs	333
9.8.4	Partially Balanced Incomplete Block Designs	335
9.8.5	Extended Block Designs	337
9.8.6	Some General Remarks	338
9.9	SYSTEMATIC BLOCK DESIGNS	340
9.9.1	Dealing with Trends	340
9.9.2	Trend-free Designs	341
9.10	EXAMPLES USING SAS®	343
9.11	EXERCISES	366
10	Latin Square Type Designs	373
10.1	INTRODUCTION AND MOTIVATION	373
10.2	LATIN SQUARE DESIGN	374
10.2.1	Definition	374
10.2.2	Transformation Sets and Randomization	376
10.2.3	Derived Linear Model	377
10.2.4	Estimation of Treatment Contrasts	380
10.2.5	Analysis of Variance	382
10.2.6	The Model under Additivity in the Broad Sense	385
10.2.7	Consequences of Nonadditivity	386
10.2.8	Investigating Nonadditivity	387
10.2.9	Miscellaneous Remarks	389
10.3	REPLICATED LATIN SQUARES	390

10.3.1	Different Scenarios for Replication	390
10.3.2	Rows and Columns Crossed with Replications	391
10.3.3	Rows Nested in and Columns Crossed with Replications	391
10.3.4	Rows and Columns Nested in Replications	392
10.3.5	Replication \times Treatment Interaction	392
10.4	LATIN RECTANGLES	393
10.5	INCOMPLETE LATIN SQUARES	394
10.6	ORTHOGONAL LATIN SQUARES	395
10.6.1	Græco-Latin Squares	395
10.6.2	Mutually Orthogonal Latin Squares	396
10.7	CHANGE-OVER DESIGNS	397
10.7.1	Two-Treatment Change-Over Design	398
10.7.2	Change-Over Designs for More than Two Treatments	401
10.7.3	Some Variations and Extensions	402
10.8	EXAMPLES USING SAS®	404
10.9	EXERCISES	414
11	Factorial Experiments: Basic Ideas	419
11.1	INTRODUCTION	419
11.2	INFERENCES FROM FACTORIAL EXPERIMENTS	420
11.3	EXPERIMENTS WITH FACTORS AT TWO LEVELS	422
11.3.1	Definition of Main Effects and Interactions	422
11.3.2	Estimation of Main Effects and Interactions	425
11.3.3	Sums of Squares for Main Effects and Interactions	426
11.4	INTERPRETATION OF EFFECTS AND INTERACTIONS	426
11.5	INTERACTIONS: A CASE STUDY	428
11.5.1	The Experiment	428
11.5.2	The Model	428
11.5.3	The Analysis	430
11.5.4	Separate Analyses	439
11.5.5	Blocking by Intrinsic Factor Only	440
11.5.6	Using the Half-normal Plot Technique	441
11.5.7	The Analysis	443
11.5.8	Summary	446
11.6	2 ⁿ FACTORIALS IN INCOMPLETE BLOCKS	446
11.6.1	2 ³ Factorial in Blocks of Size 4	446
11.6.2	2 ³ Factorial in Blocks of Size 2	447
11.6.3	Partial Confounding	449
11.7	FRACTIONS OF 2 ⁿ FACTORIALS	451
11.7.1	Rationale for Fractional Replication	451
11.7.2	1/2 Fraction of the 2 ³ Factorial	454
11.7.3	The Alias Structure	454
11.7.4	1/4 Fraction of the 2 ⁸ Factorial	456
11.7.5	Systems of Confounding for Fractional Factorials	457

11.8	ORTHOGONAL MAIN EFFECT PLANS FOR 2^n FACTORIALS . . .	462
11.9	EXPERIMENTS WITH FACTORS AT THREE LEVELS	464
11.9.1	The 3^2 Factorial	465
11.9.2	Extensions	468
11.9.3	Formal Definition of Main Effects and Interactions	468
11.9.4	Systems of Confounding for the 3^n Factorial	470
11.9.5	Fractions of 3^n Factorials	472
11.9.6	Highly Fractionated 3^n Factorials	475
11.9.7	Systems of Confounding for Fractions of 3^n Factorials	475
11.10	FACTORS AT TWO AND THREE LEVELS	476
11.10.1	Asymmetrical Factorial Experiments	476
11.10.2	Confounding in $2^m \times 3^n$ Factorials	477
	Blocks of Size 18:	478
	Blocks of Size 12:	478
	Blocks of Size 9:	478
	Blocks of Size 6:	478
	Blocks of Size 4:	478
11.10.3	Fractions of $2^m \times 3^n$ Factorials	479
11.11	EXAMPLES USING SAS®	481
11.12	EXERCISES	492
12	Response Surface Designs	497
12.1	INTRODUCTION	497
12.2	FORMULATION OF THE PROBLEM	498
12.3	FIRST-ORDER MODELS AND DESIGNS	500
12.3.1	First-Order Regression Model	500
12.3.2	Least Squares Analysis	500
12.3.3	Alternative Designs	503
12.4	SECOND-ORDER MODELS AND DESIGNS	504
12.4.1	Second-Order Linear Regression	504
12.4.2	Possible Designs	505
12.4.3	Central Composite Designs	506
12.4.4	Blocking in Central Composite Designs	507
12.4.5	Box-Behnken Designs	509
12.4.6	Hard-to-Change versus Easy-to-Change Factors	511
12.5	INTEGRATED MEAN SQUARED ERROR DESIGNS	513
12.5.1	Variance and Bias for the One-Factor Case	514
12.5.2	Choice of Design	517
12.6	SEARCHING FOR AN OPTIMUM	518
12.7	EXPERIMENTS WITH MIXTURES	519
12.7.1	Defining the Problem	519
12.7.2	Simplex-Lattice Designs	520
12.7.3	Simplex-Centroid Designs	521
12.7.4	Axial Designs	521
12.7.5	Canonical Polynomials	521

12.7.6 Including Process Variables	523
12.8 EXAMPLES USING SAS®	523
12.9 EXERCISES	531
13 Split-Plot Type Designs	533
13.1 INTRODUCTION	533
13.2 SIMPLE SPLIT-PLOT DESIGN	534
13.2.1 Superimposing Two Randomized Complete Block Designs . .	534
13.2.2 Derived Linear Model	537
13.2.3 Testing of Hypotheses	538
13.2.4 Estimating Treatment Contrasts	539
13.2.5 Testing Hypotheses about Treatment Contrasts	542
13.3 RELATIVE EFFICIENCY OF SPLIT-PLOT DESIGN	543
13.4 OTHER FORMS OF SPLIT-PLOT DESIGNS	544
13.4.1 SPD(CRD, RCBD)	545
13.4.2 Split-Plot Design in Time	545
13.4.3 SPD(CRD, LSD)	547
13.4.4 SPD(LSD, RCBD)	548
13.4.5 SPD(CRD, IBD)	549
13.4.6 SPD(GRBD, RCBD)	550
13.4.7 SPD(GRBD, IBD)	552
13.4.8 SPD(IBD, RCBD)	553
13.4.9 SPD(RCBD, GRBD)	554
13.4.10 Summary	555
13.5 SPLIT-BLOCK DESIGN	555
13.5.1 The Layout	555
13.5.2 Linear Model and ANOVA	557
13.5.3 Estimating Treatment Contrasts	557
13.6 SPLIT-SPLIT-PLOT DESIGN	560
13.7 EXAMPLES USING SAS®	562
13.8 EXERCISES	569
14 Designs with Repeated Measures	573
14.1 INTRODUCTION	573
14.2 METHODS FOR ANALYZING REPEATED MEASURES DATA . .	574
14.2.1 Comparisons at Separate Time Points	574
14.2.2 Use of Summary Measures	575
14.2.3 Trend Analysis	575
14.2.4 The ANOVA Method	577
14.2.5 Mixed Model Analysis	578
14.3 EXAMPLES USING SAS®	580
14.4 EXERCISES	593
Epilogue	595

CONTENTS

xv

Bibliography

597

Abbreviations

613

Author Index

615

Subject Index

619

This Page Intentionally Left Blank

Preface to the Second Edition

Imagine the following telephone conversation between a statistician (S) and a research scientist (R). R: "Hello, Mr. Stat, I wonder whether you have just a minute for a quick statistical question." S: "Usually I do not do statistical consulting over the phone, but let me see if I can help you. What is the problem?" R: "We are developing new growth media for industrial producers for growing flower plants. We have three such media and we use them with four flower varieties. We have five replications for each combination of medium and flower. We have analyzed the data as a 3×4 two-way classification with five observations per cell. But my graduate assistant has talked to one of your students and he is now confused about the validity of this analysis. I just want you to confirm that we have done the right thing." S: "Well, I do not know." R: "What do you mean, you do not know? You are the expert!" S: "I really need to know more about how you performed the experiment. For example, how did you prepare the media that you used in the individual pots? I assume that you grow the flowers in pots in the greenhouse." R: "Yes, that is right. My graduate assistant simply mixed each medium in a big container, which we then put in the individual pots." S: "That may be a problem, because now you may not have any replication." R: "What do you mean, we have no replication? I just told you that we have five replications." S: "Yes, but . . . I think it would be best if you would come to my office for me to explain this to you and to take a closer look at your experiment." R: "But we have already submitted the paper for publication." S: "Then why don't you come when you get the reviews back." Silence. R: "Yes, thank you. I'll do that."

This is, of course, just a fictitious conversation. But many consulting statisticians have had similar conversations. The aim of this book is to help statisticians as well as research scientists not only to better understand each other but also to obtain a better understanding of the intricacies of designing and analyzing experiments. It is our hope that this can be achieved by using the book as a textbook as well as a reference book.

Having used the first edition for several years as a textbook in an MS level class for statistics students and well-qualified graduate students from other fields relying heavily on experimental research, I have gained valuable insight into the needs of both types of students. This has led to some changes and enhancements in the second edition without giving up the general flavor and philosophy of the book.

Although some readers may feel the book is too theoretical, we strongly believe that these developments are necessary to understand the basic ideas and principles of experimental design, to enable students and researchers to pursue ideas of designing experiments not covered in this book, and to lay the foundation for even more theo-

retical work as covered, for example, in Volume 2: *Advanced Experimental Design*. For the non-statistics student it is not always necessary to understand all the details leading to important results as long as they develop a certain feel for these results and appreciate their role and importance in the overall picture. A skillful teacher will be able to accomplish these aims without compromising the rigor of the development of the material.

Having said this, I have tried to make the second edition more user-friendly by also emphasizing the practical aspects of designing and analyzing experiments. I have considerably expanded Chapter 2 with further discussion of the planning aspects of setting up an experiment and giving heuristic arguments why the various steps are so important for a successful experiment. This should appeal to both consulting statisticians and research scientists.

Other major changes involve the development in Chapter 9, which I consider to be one of the most important chapters in the book because of the introduction of the notion of blocking. I spend a great deal of time discussing the different types of blocking factors and their importance in the overall scheme of setting up, analyzing, and drawing inference using various forms of block designs. These ideas are then carried over to Chapter 11, which introduces the basic concepts of factorial treatment structure and design. I have included a case study, based on an invited presentation at a meeting of the American Society for Horticultural Science, which discusses in some detail the role, analysis, and interpretation of various forms of interactions.

The discussion about repeated measures has been moved to a separate chapter (Chapter 14) to give it more emphasis, as this type of experimentation occurs quite often. I explain how repeated measures can be paired with any error-control design, how this leads to a split-plot type structure of the experiment, and how and why the analysis differs from that of a split-plot experiment.

Finally, I have added to most chapters numerical examples using the statistical software package SAS® (SAS Institute, Inc. 2002–2003) as a tool to analyze the data. This should not be considered to be a tutorial in SAS, but it should provide some help to readers of this book about how to analyze similar data from their experiments and to relate such analyses to the developments, in particular ANOVA tables, given in the book. In order to preserve space I have omitted some information provided in the usual SAS output. Also, I should mention that the data are not real, even though some of the experiments described are, as research scientists are generally not willing to share their original raw data. The results presented should, therefore, not be interpreted as findings in a given subject matter area, but rather as illustrations of statistical procedures useful for analyzing such data. For readers who do not have access to SAS or who prefer to use other statistical software, the examples should provide some help in setting up the analyses in their environment. In addition to using SAS as a tool for the analysis from designed experiments I also show how SAS can be used for randomization procedures and for constructing certain types of factorial designs.

I hope that the changes and enhancements in the second edition will prove useful to students, teachers, and researchers. For those who seek a deeper understanding and further developments of the material presented here I provide references to chapters and sections in Volume 2 indicated, for short, by II.xx or II.xx.yy, respectively.

An FTP (ftp://ftp.wiley.com/public/sci_tech_med/design_experiments/) for this book

will be maintained by wiley.com, which will also contain additional exercises and solutions to selected exercises.

During the process of thinking about and completing this revision I have received help from several people. I would like to thank my students and colleagues for pointing out errors in the first edition and for making suggestions for changes. I am grateful to Yoon Kim and Ayca Ozol-Godfrey for their help with some computational and graphical aspects in Chapters 6 and 11. It is difficult to find the right words to express my profound gratitude to Linda Breeding for her tireless and skillful efforts in producing the camera-ready manuscript. This has been a monumental and difficult job, and even during times of despair she found a way to carry on until the work was completed. Nobody could have done it better. Thank you, Linda. I also would like to thank Jonathan Duggins, Amy Hendrickson, and Scotland Leman for their expert advice and help with LaTeX.

Finally, I would like to dedicate this edition to the memory of my co-author and mentor, Oscar Kempthorne, for his many important contributions to the philosophy, theory, practice, and teaching of experimental design (for a bibliography, see Hinkelmann, 2001).

KLAUS HINKELMANN

This Page Intentionally Left Blank

Preface to the First Edition

The subject of the design of experiments has been built up largely by two men, R. A. Fisher and F. Yates. The contributions of R. A. Fisher to mathematical statistics form a major portion of the subject as we now know it. His contributions to the logic of the scientific method and of experimentation are no less outstanding, and his book *The Design of Experiments* will be a classic of statistical literature. The contributions of F. Yates to the field of the design of experiments are such that nearly all the complex designs of value were first put forward by him in a series of papers since 1932. Both Fisher and Yates have also made indirect contributions through the staff of the statistical department of Rothamsted Experimental Station, since its founding in 1920. It is not surprising that the contributions originated from Rothamsted, because Rothamsted was probably the first place in the world to incorporate a statistical department as a regular part of its research staff, and the design of experiments is a subject that must grow through stimulation by the needs of the experimental sciences.

This quotation from the preface of *Design and Analysis of Experiments* by Kempthorne (1952) affirms our recognition of the enormous and path-breaking contributions made by these two men to the field of experimental design and experimentation in general. Even though most of their ideas originated in connection with agricultural or genetic experiments, the resulting principles and designs have found wide applicability in all areas of scientific investigations as well as in many areas of industrial production and development.

Because of the widespread use and increasing importance of experimental design, it is essential that students and users obtain a firm understanding of the philosophical basis and of the principles of experimental design as well as a broad knowledge of available designs together with their assumptions, their construction, their applicability, and their analysis. These topics then are the subject of this book which will appear in two volumes.

Volume I is a general introduction to the subject laying the foundation for the development of various aspects of experimental design. In it we describe and discuss many of the commonly used designs and their analyses. We return to some of these designs and introduce other designs in Volume II at a more technical and mathematically more advanced level.

With respect to the present volume, Chapters 1 through 5 describe in some detail the philosophical foundation and the mathematical-statistical framework for our approach to the discussion of experimental design. We put the notion of and the necessity for intervention studies, the main topic of this book, squarely into the context of the

scientific method. We develop and draw a sharp distinction between observational and intervention studies, a theme to which we return at various places throughout the book, in particular in connection with the analysis of data. Much of the analysis is based on the theory of linear models. A thorough discussion of linear models theory is given in Chapter 4. Our major aim here is to provide the reader with the basic tools to understand and develop the analysis of data from intervention studies of the sort discussed in this book.

Although linear models play a fundamental role we stress the fact that they do not exist in and of themselves but that they evolve from very basic principles and in the context of the experimental situation at hand. Indeed, in Chapter 2 we argue that many facets are involved in advancing from a research idea or question to a designed experiment which permits the investigator to draw valid conclusions. Some of these facets are of a statistical nature such as developing an appropriate experimental design, developing an appropriate model, and carrying out an appropriate analysis, and they are the subject of this book. But it is important, we assert, to always keep in mind that statisticians and subject-matter scientists have to combine their knowledge to develop an experimental protocol according to sound principles of both fields.

We have alluded earlier to the impact that R. A. Fisher had on the development of the field of experimental design. One of his contributions concerning the design of experiments is the use of randomization. In Chapter 5, as well as in following chapters, we discuss the general idea and then apply it to specific designs. It forms the basis of the analysis for all intervention studies.

Beginning with Chapter 6 we develop from first principles various error-control designs. We start with the completely randomized design (Chapter 6) as the simplest form of error-control design and then move on to more complex error-control designs such as randomized block designs (Chapter 9), Latin square type designs (Chapter 10) and split-plot type designs (Chapter 13). For each design we derive linear models and the associated analyses, mainly in the form of analyses of variance. Other forms of analysis such as estimating and testing treatment contrasts are dealt with in Chapter 7. And further reduction of experimental error through the use of supplementary information is described in Chapter 8.

The notion of treatment design is introduced in Chapter 11 when we discuss factorial experiments. Particular attention is paid to experiments involving factors with two and three levels. This serves as an introduction to the vast opportunities and techniques that are available for such type of experimentation. We emphasize in particular how treatment designs can be combined with or embedded in error-control designs in the form of systems of confounding.

In Chapter 12 we touch briefly on a different form of experiment designs: response surface and mixture designs. It serves mainly to point out the difference between comparative and absolute experiments, but it also serves to show how error-control designs and treatment designs can be applied towards the construction of response surface designs.

Although many experiments can be conducted using the designs discussed here and in Volume II, there are many others for which special designs need to be constructed. It is our aim here to lay the foundation for such work by discussing in detail the major principles of experimental design, such as randomization, blocking (in particular in-

complete blocks), the Latin square principle, the split-unit principle, and the notion of factorial treatment structure.

The relationship of the present two volumes to the book *Design and Analysis of Experiments* by O. Kempthorne published in 1952 merits some discussion. Very much is common. We have felt it absolutely necessary to add a long chapter on the process of science, discussing our perception of observation theory in science, the role of experiments, the role of data analysis, and the introduction of ideas of probability as related to relative frequency in a defined population of repetitions. The presentation of least squares and the general linear hypothesis needed large improvement. We have based most of the data analysis and inference on randomization analysis, expanding and formalizing the presentation. The remainder of the presentation in the present two volumes is a considerable expansion and rearrangement of standard material of the 1952 book taking many of the developments during the last 40 years into account.

The organization and presentation of the material has evolved over a number of years of teaching the subject to graduate students in statistics. Volume I is intended as a textbook for a one-semester course for first year graduate students. To make the course effective, the students should have been exposed to a fairly rigorous course in statistical methods. They should be familiar with the basic principles of statistical inference and with the rudimentary ideas of analysis of variance and regression, i.e., they should have some understanding of and appreciation for linear models and their role in statistical inference.

The book contains more material than can be taught reasonably in one semester, and hence a selection of topics will have to be made. This will depend to some extent on the students' background and preparation. One suggestion is to skip some details and omit certain parts in individual chapters. Another is to omit much of Chapter 4 and Chapter 7 and omit all of Chapter 12 (at Virginia Tech, for example, there exists a concurrent course in the theory of linear models covering the material in Chapter 4, most of the material in Chapter 7 will have been covered in a course on applied statistics, and there exists a separate course for response surface designs). At any rate, Chapters 6 through 13 are fairly self-contained except that frequent reference is made to results in Chapter 4 for a better understanding of the underlying principles.

The reader will notice that no numerical examples are given throughout the text. It is assumed that the reader is familiar with mathematical notation and does not have any difficulty with reading and handling formulas. There is no emphasis at all on calculations. Instead, we provide some guidance on how to use available statistical software. This attempt is, however, rather limited in that we refer only to SAS as an example of available software packages, and even that is by necessity not complete.

A thorough knowledge of the material in Volume I is a prerequisite for understanding Volume II. As mentioned before, the presentation of the material in Volume II is more technical and hence suited for a more advanced course in experimental design. It contains more than enough material for another one-semester course. In addition, Volume II is intended to serve as a reference book on many topics in experimental design.

Many people have contributed to this book in different ways. Foremost among them are our students who have been exposed to this material over the years. We thank them for their questions and comments. K. H. would like to thank Virginia Tech for granting

him study-research leaves and the Departments of Statistics at the University of New South Wales and Iowa State University for providing him with support, facilities, and a congenial atmosphere in which to work. O. K. is grateful to Iowa State University for providing more than 40 years of stimulating environment and association with many graduate students of high ability. We thank Yoon Kim and Sungsue Rheem for help with the simulations for the randomization analyses, and Markus Hüttmann and Sandra Schläfke for extensive help with the preparation of the index. And finally, we express our deep appreciation and gratitude to Linda Breeding and Ginger Wenzlik for their expert typing and word-processing.

KLAUS HINKELMANN
OSCAR KEMPTHORNE

CHAPTER 1

The Processes of Science

1.1 INTRODUCTION

In order to understand the role of statistics, generally, and the role of design of experiments in particular, it is useful to attempt to characterize the processes of science and technology. All science and technology starts with questions or problems. The grand aim is to develop a model which will describe adequately, that is, accurately, the past, present, and future of the universe. Obviously, if we are to describe the future, we must have a model that incorporates development over time—that is, a dynamic model, and a model that predicts what alterations will be brought about by interventional acts, such as drug therapy, reducing money supply, or supplying a nation with armaments, to give widely disparate examples.

1.1.1 Observations in Science

The foundation of all science is, obviously, observation. This, which we all do every waking minute of our lives, would seem to be a very simple matter, with a logic that is entirely clear. It is not clear from several points of view. Curiously enough, it is not discussed, it seems by philosophers of knowledge. It is obvious that animals make observations—all one has to do is to try to catch a rabbit. It observes that it is being chased and takes evasive action. This is, presumably, an instinct bred into rabbits by the evolutionary process. In science, a reaction to a portion of the world is an observation only if that reaction can be recorded, perhaps only in memory, or better, of course, by actual physical recording. To do this requires a language and descriptive terms. It is necessary that an observation can be described in terms that have some meaning to others. The development of a language for this purpose, a language that is effective, is a process of science that continues. We need only look at the development of the language of biology. This field is full of names of things, and indeed, one of the great difficulties of the field is to learn the naming that has been developed in the past, a task that becomes more and more difficult as processes of observation are being developed, one can almost say, day by day. Many parts of the journal *Science* of today are unreadable except by experts and would be unreadable for the experts of decades

ago. The development of this type of descriptive language proceeds with care, and with the discipline of the area of study. A descriptive term does not receive validation until it is agreed on and can be confirmed by any observer who follows the prescribed protocol of observation and has been educated in the use of the descriptive terms. This is no more than a cliché in physical and biological science and one might be led to the view that it is not worth stating. But when we turn to any aspect of human mental status or mental behavior, the “obvious” cliché becomes critical. One merely has to look at the nosology that occurs in psychiatry to see the problem. This is not to imply that workers in that area are dolts—the area is remarkably difficult because of the problem of validation of observation.

A second point about observation is that it is by its very nature incomplete. One observes, one says, a robin outside one’s window. Humanity uses this mode of expression and it has served it well. But one does not observe the whole of the phenomenon. Just recall the commonplace interchange. Person A says “I see a robin.” Person B says, “Yes, I see it too. Did you notice that it has a gray bar on its wing tips?” Person A says, “No, I did not notice that, but now that you mention it, I do see it.” Person A’s observation was incomplete relative to B’s observation. Obviously, there can be person C, who sees more. Also, obviously, observation is not an innate ability; it is one which may require high “professional” training—even in areas that use no more than the ordinary unaided human eye. For the naturalist of the sixteenth century, for the ordinary citizen naturalist, and for the person who has received two years of training, observation is not at all the same. If we adjoin the obvious massive development of observational processes, with physical devices, for instance, observing in infrared light, observing with an electron microscope, and so on, there is not an elemental activity which we can call “making observation.”

Another aspect, which is much more subtle, is that the process of observation may or may not have an effect on what is being observed, with the elementary consequence that one simply cannot observe the status of an object of observation. There are, of course, elementary techniques for combating this, as in the use of blinds to observe birds, or of walls that have one-way vision. But when one considers observing, or trying to observe, the mental state of a human and leaving aside the possibility of observing what one thinks to be physical correlates of mental status, one has to talk to the human and ask questions and then it is not at all clear what the status of ensuing conversation by the human being observed is. Turning aside from an obviously fantastically difficult area, we saw a revolution in physics at the beginning of the twentieth century with the realization that one could not look at a particle except by shooting another particle at it and getting a collision. This type of situation led to the famous Heisenberg uncertainty principle in an area for which it was thought previously that one could observe without interfering with the object observed. This phenomenon has huge consequences as any modern physicist knows. It has, also, huge consequences with respect to epistemology.

1.1.2 Two Types of Observations

We leave this discussion. For the purposes of our discussion here, we assume that there is a validated process of observation that has no effect on the object being observed. We

must, however, discuss a major point. There are fundamentally, it seems, two types of observation. The first consists of placing an observation of an object of observation in a class: for instance, the flower being observed is pink or has pinnate leaves. In most circumstances, there is no doubt of the recorded observation (though one can be doubtful, e.g., on a color designation). In other cases, the result of the observation is uncertain; we merely have to imagine being given sequentially with repetition unknown to the observer of a set of colored blocks that do not have strongly distinguishable colors. One will find that one's observation of a block will give different results in repetitions, over which one is fairly sure that color has not changed. In such cases, one has no recourse but to use a probability model to the effect that, in repetitions that are unconnected in a known way, there will be a frequency distribution of observational outcomes. We shall not be concerned with this at present. The second type of observation, which permeates quantitative science, is the measurement of a numerical magnitude, for example, the weight of a piece of rock, which one is confident does not change. In this case, there is always an error and an imprecision of measurement. The nature of the error and of the imprecision is again representable by some frequency distribution of results. This type of problem permeates, of course, the physical sciences, and increasingly so, as the sought after observation, such as weight, becomes smaller and smaller.

We hope that we have given a useful discussion of observation, though elementary and potentially highly obscure at a philosophical level. We take comfort in the fact that even if the process of observation is quite unclear (as it is at a fundamental level), the world of science is permeated with interpersonally validated observation.

1.1.3 From Observation to Law

Our writing here is aimed at constructing a useful model of what happens in science. It seems clear that the beginning of science is observation and description. It also seems clear that this is still a critical feature of science. This observation process requires no theory. It is interesting in this connection, to look at what Darwin (1809–1882) did in the voyage of the *Beagle*. He was, for his age, a very remarkable observer. The point of expressing the above views is that it is sometimes said that the mere collection of observations has to be based on a theory, a concept to which we shall turn. If one wishes to state that even the simplest observation is based on the informal theory that one's observation process obtained an attribute of what is observed and not of the observer, one cannot object. Apart from this, much observation has been generated not by any theory but by curiosity, an attribute that one observes in animals. It is true, of course, that, very often, observation is initiated by a question or by a problem situation. One could say that curiosity is the result of a question, but this seems to be mere playing of a verbal game. Obviously, our presentation is Baconian and we quote his first Aphorism (Spedding et al., Vol. I, 1861, p. 241):

Homo, Naturæ minister et interpres, tantum facit et intelligit quantum de Naturæ ordine re vel mente observaverit, nec amplius scit aut potest.

for in its translation (Spedding et al., Vol. VIII, 1863, p. 67):

Man, being the servant and interpreter of Nature, can do and understand so much and so much only as he has observed in fact or thought of the course of nature:

beyond this he neither knows anything nor can do anything.

However, we cannot accept the full Baconian prescription as given by his successive aphorisms.

It is obvious that science does not consist merely of a collection of, shall we say, interpersonally validated observations. What comes next? Rather clearly, it is the organization of such observations—let us call them facts, into sets of related facts. Let us suppose that our observations are categorical. We observe trees. This consists of noting, with our developed language, that there are trees that keep their leaves through the winter and those that do not. We deliberately take simple examples that even the proverbial man-in-the-street can appreciate. So this part of our observation places the objects of observation into one of an exhaustible polytomy. Let us call the classes of one polytomy $\alpha_1, \alpha_2, \dots, \alpha_r$ and of the second, $\beta_1, \beta_2, \dots, \beta_s$. We look at our observations and we see that in our observations every object which was α_3 is β_2 . Obviously, a generalization is suggested: every object that is α_3 is β_2 . We have a suggestion of a “law.” The word “law” is used in our language in many senses and even in science it is used in at least two senses. A “law” states that something *must* occur. The Creator has decreed so. This is one sense. Another sense, which is really quite different, is that a law is an empirical generalization. A hoary and false example is: I have seen 10 swans and they were all white. So I infer (falsely) the “law” that all swans are white. This example leads, of course, into the problem of induction, on which libraries of books have been written, *without resolution*.

We then see a very curious thing happening, the development of a theory. From the “law” obtained as a suggested empirical generalization, we convert our generalization into a “law” of Nature, something that must necessarily be the case. When we do this, we are beginning to make a theory. This is, however, just one part of the construction of a theory. It is the absence of a role for theory that has been the main criticism of the prescription of Bacon (1560–1626).

It is informative, here, to bring in the work of Kepler (1571–1630). The observations were the positions of planets at different times of the year. The contribution of Kepler was to analyze the data and to show that the path of each planet was an ellipse, with the sun as focus, and other aspects that are given in his famous three laws. It is also informative to recall the work of Mendel (1822–1884) in biology. The crossing of types X and Y gave offspring of type Z, say, and then the crossing of these offspring gave an array of offspring which had the *appearance* that $\frac{1}{4}$ were X, $\frac{1}{2}$ were Z, and $\frac{1}{4}$ were Y. Interestingly, this appears to be the first case of occurrence of a validated probability model in science (apart from mere gambling).

In the one case, we have Kepler’s laws and, in the other case, Mendel’s laws. Now, we have to raise the hard question. Are these laws as empirical generalizations or are these laws that tell us what must happen? Are they built into Nature by the Creators? Our answer is obvious, we think: they are *merely* empirical generalizations.

How then do we get to a theory? The process is rather simple, though rarely expounded in our experience. At first, we have what may be called “naive” laws, merely empirical generalizations. But we want an explanation. We are in a morass and we shall have to discuss the idea of “explanation.”

1.2 DEVELOPMENT OF THEORY

The suggestion of Bacon that all we need to do is make observations has been rather uniformly criticized in the ensuing four centuries or so. We suggest that the criticism has not been entirely justified. One question is: what observations should one make? Obviously, we could suggest that the way to understand the universe is to have millions of humans observing—observing, only observing. It is obvious, indeed, that such a program would lead to an incredible mass of observational facts. The first missing ingredient in the Baconian prescription is given by the question: “What are we to do with all the facts that are obtained?” This question is interesting to the field of statistics, and eventually, to the whole of science, because it tells us that we have to do “data analysis.” It is interesting and curious that this is a term used to denote an activity that has been pursued by humans from the beginning of time, but which has been popularized since the 1960s in all discussions of statistics.

It is surely pretentious to think that one can encapsulate the efforts of humanity to understand the universe in a few printed pages. But it is useful, we think, to attempt to present a broad picture that captures essential features of the human efforts that have been made. Interestingly, this is, of course, a problem of data analysis of its own. We can look over the history of science. None of us can do this completely, but perhaps we can see a general pattern which is not misleading.

1.2.1 The Basic Syllogism

The beginning is surely observation of entities that have some degree of permanence in time, usually entities that one can, so to speak “hold in one’s hand,” literally or metaphorically. These are looked at and classified. This is just Aristotelian classification. From this came laws as empirical generalizations such as “All A are B,” or “All entities which have attributes α and β have attribute γ .” It is interesting that this led to the basic syllogism: (i) All entities with α have β ; (ii) entity E has α ; therefore, (iii) entity E has β . We find this presented as a mode of deduction, and this matter needs discussion. This syllogism is used widely and essentially in mathematical reasoning, and without it, the possibility of the sort of mathematics we do would be impossible. For instance, every triangle has the property that the sum of its interior angles is 180° : here is a triangle; therefore, the sum of its interior angles is 180° . From where do we get the first part of the syllogism? The answer is simple. We prove it! But then we have to ask: “What do you mean by that?” What does it mean to say: “We prove it”? The answer is given in simple and not misleading form. We have defined “triangle.” We have developed modes of deduction that we *accept* as constituting proof. This whole process is very subtle. Just how subtle it is can be seen from the developments of mathematics of the past two centuries or so. We see students in high school trying to write proofs in geometry. We see ourselves writing proofs that we judge to be complete. But we later see that our proofs are incorrect or incomplete. We see in the history of mathematics incorrect proofs by great mathematicians. We see proofs in which a questionable syllogism has been used with total unawareness that it has been used. The curious outcome of this phenomenon is that a proof of a mathematical theorem is a sequence of statements, in mathematical form, developed from axioms that

are unquestionable, that the world of mathematics accepts as constituting proof. The purported proofs have been examined by thousands of mathematicians and found to be convincing. This should not be taken to be derogative and pejorative. The world of non-mathematicians should know that there is considerable controversy at the foundations of mathematics, a controversy that has arisen *only* in the past century or so. What are the axioms that we are to accept as indubitable?

Our interest in the basic syllogism is not with respect to its use in mathematics, but its use in describing and explaining the real world—the world we can observe. It seems entirely clear that the use of the syllogism in this context is totally questionable. It is questionable from the point that it is empty. If we know that all A's are B, and we know that X is an A, we are allowed to *deduce* "Therefore X is B." But this is an empty deduction, because with respect to the real observable world, we cannot use as a premise that all A's are B without having assured ourselves that X, which is an A, is B. We read texts on logic and we find the standard example:

All men are mortal.
Socrates is a man.
Therefore, Socrates is mortal.

Can we use this in *deduction* about the real world? The problem is, of course, the validity of the first statement—the premise. As we have said, by accepting that all men are mortal, that Socrates is a man, we have accepted the so-called conclusion. From one point of view, we are just playing a word game, and are hoping to impress our reader by using the very heavy word "therefore." A curious example was given in the popular press recently. Consider the sequence of statements:

All babies are nurtured in a uterus before birth.
Individual X is a baby.
Therefore X was nurtured in a uterus.

The point of the example is that the uterus of a mother had been removed some years before.

Surely one cannot say that the basic syllogism is nonsense. Without it, no science would be possible. What then is going on? The answer to this question is very simple *in form*. If the premise is to be useful, it must be established independently of the individual use of it; in other words, we must have the knowledge that all men are mortal, without having observed that Socrates is mortal. The upshot is then obvious; to establish the premise in such a way as to be useful in the syllogism, we must use induction. We can do no more than say: We have examined many humans and found that everyone of them was mortal. So we induce that mortality is a universal property of the class "humans." Then if we see that Socrates is in the class "humans," it must be the case that Socrates is mortal.

1.2.2 Induction, Deduction, and Hypothesis

The punch line of this stream of thinking is that the use of the basic syllogism *as a tool of science* is based on induction. Or, with perhaps a harsh mode of statement, deduction as applied to the real world is totally ineffective without the establishing of the premise

by induction. Bacon in his Aphorism XIV stated (Spedding et al., Vol. VIII, 1863, p. 70): "Our only hope therefore lies in a true induction." So we have to try to say something reasonable about induction. This is a very difficult task; we merely have to read the many volumes on the topic. To exemplify the difficulties, we quote Bertrand Russell (1959):

But there is a much more general problem involved here, which has continued to bedevil logicians to the present day. The difficulty is, roughly, that somehow people feel induction is not after all as respectable as it ought to be. Therefore it must be justified. But this would seem to lead to an insidious dilemma that is not always recognized. For justification is a matter of deductive logic. It cannot itself be inductive if induction is what must be justified. As for deduction itself, this no one feels compelled to justify, it has been respectable from time immemorial. Perhaps the only way is to let induction be different without seeking to tie it to deductive apologies.

This statement deserves comment. It tells us, clearly, that Russell (1872–1970), surely one of the ten or so finest minds of the twentieth century, does not help us. It tells us that Russell cannot help us with the problem of understanding and carrying on science, because it is obvious, we assert, that one of the primary bases of science is induction.

The foremost philosopher of science, perhaps, was C. S. Peirce (1839–1914) (see Gallie, 1966). We cannot, here, give our detailed understanding (which may be fallacious) of his ideas. Peirce distinguished three types of inference: deduction, induction, and hypothesis. The third he preferred to call "abduction", which, it seems, is a method of testing rather than of developing knowledge. Workers in statistics will have no difficulty in appreciating this third type: a considerable portion of statistical theory and practice is the testing of statistical models. This, of course, entertains the possibility that a model or a theory can be shown to be false. The essential feature of science is that its theories can be "falsified." This view is supported strongly, it appears, by the philosopher of science, Karl Popper (1902–1994). The only problem we see with his writings is an absence of an approach to methods of falsifying hypotheses or models. If we have a universal, "All A's are B," how are we to falsify it? Rather obviously, the only thing we can do is to continue to examine the A's that we meet and see if they are B. A single occurrence of A and not B falsifies the universal. Suppose, however, that we have observed 100 A's and find that they are all B. Does this justify the universal? Obviously not. It does, of course, suggest it. Can we quantify strength of support for the universal? Obviously, we should feel more confident if we found the occurrence with 100 A's rather than 10 A's. We shall not pursue this discussion except to state our view that this problem can be addressed only by making an assumption of randomness, which must be questioned, followed by tests of significance and tests of hypotheses. These are hypothesis "falsification" procedures. If our hypothesis is that an unknown, which is a constant in a theory takes a certain value or lies in a certain range we shall again use statistical tests, and associated statistical intervals.

No theory put forward so far, even in physics, the so-called "Queen of Science", has withstood the test of falsification. We do not bother to substantiate this; we merely advise the reader of this exposition to look over the sequence of theories. A facet of this must be discussed. Even though a theory, for example, the theory of gravitation or the theory of electricity and magnetism of some past period, has been shown to be

false, that theory may be excellent for predicting a wide variety of outcomes of circumstances. Our ordinary living is based, with the use of electricity, on what may be called the classical theory of electricity and magnetism, and in such application this theory is excellent, and obviously so. This tells us something that is highly significant. We cannot talk about a theory being absolutely true. We can only talk about a theory being true in a given context of application. Obviously, we have many such theories which we use every day, in devising, for instance, the “gadgets,” heating systems, cooling systems, transportation systems, sending man to the moon, etc., and in the nutritional “theories” that we use for plant, animal and human nutrition, etc., and the medical theories we use, such as those to cure deadly diseases, such as syphilis or gonorrhea, etc., or to palliate chronic diseases, such as diabetes.

We do not have the time or the ability to pursue this line of thought. However, it leaves us with the view, that we hold rather firmly that the question of whether a theory is true, unconditionally, is *not* a well-formed question. We have to ask if the application of a given theory to a specified set of circumstances gives a prediction that is verified to be correct.

1.3 THE NATURE AND ROLE OF THEORY IN SCIENCE

We read writings, which we shall not cite, which take the position that there cannot be science without scientific theory. We may mention, however, the writings of Poincaré (1854–1912) and the writings of Popper as indicating at least a strong tendency toward this view. We shall first exposit our opinion that this view is wrong. It is wrong for the very simple reason that there are varieties of science.

It is absurd for any writer to claim that he or she can classify science into well-defined disjoint activities. However, any writer who pretends, that is, claims, to write about science in the broad sense must make an attempt and must recognize that there are, indeed, partially disjoint activities.

1.3.1 What Is Science?

A century ago, with some exceptions, some of which we shall mention, science was thought to consist of physics. Perhaps chemistry could be admitted to the domain of science, if only because much of it is based on physics. This view, we believe, persisted and still persists in the writings of philosophers of science. We shall not attempt to give our basis for this perception. What were the exceptions? Rather obviously, one had to admit that biology is a branch of science. As regards agriculture, it is obvious, *being ironical*, that is a problem for farmers, not for science. But should one admit psychology, sociology, economics (the so-called “dismal science”), political science, demography (and all related problems, such as nature and amount of employment, cost of living, etc.), education, child development, ecology, traffic, and so on to the august realm of science? Our view is that we should do so. Furthermore, any exposition of philosophy of science that does not give this status to the areas of investigation

mentioned and to others that could be listed, should be regarded as being so defective as not to merit deep acceptance.

It is useful, perhaps, to give some perceptions on the origin and history of the limitation of science in the way indicated. One has to go back to the Greeks, for whom science is what one knows, or science is what is *true*. This leads, of course, to the question of “What is truth?” This question has, obviously plagued humanity “since time began,” and it is clearly impossible to address this question in all its depths, for many reasons, including that of competence of the writers.

To cut a long story short, and, hopefully, not to do rank injustice to the thinkers of the past, the basic idea of proving something, that is, proving a proposition to be true, is to take as true certain axioms and then deduce the proposition from those axioms by Aristotelian logic. An early formulation of this was the process of Descartes (1596–1650), whose prescription was to subject every proposition to extreme doubt. As a result of this process, one would reach certain propositions that cannot be doubted. One would, then, have a basis for a deductive argument. The problem with this prescription is, obviously, that a process of extreme doubting will lead to nothing certain. For Descartes, the first unquestionable proposition was “*Cogito ergo sum*”—“I think, therefore I am.” Whether we can accept this translation with the present-day meanings of words is not at all clear. However, it is surely the case that this, as a basic proposition, has been questioned severely over the centuries, most recently by Sartre (1905–1980). The whole history of philosophy since Descartes has been very tangled. Certainly, highly significant thinkers were the so-called British empiricists: Locke (1632–1704), and, especially, Hume (1711–1776), who, it seems, was the first to pose the problem of induction. If we continue the development, we come to Kant (1724–1804), who had two highly significant ideas. One is that behind the world of phenomena there is a world of noumena, about which we can know nothing. If this is a correct, even if brutally short, characterization, the idea is remarkably modern. A second Kantian idea is that there are two types of truth: a priori analytic truth, which is true by virtue of language, e.g., “I am the father of my son,” and a priori synthetic truth. There can be no doubt about the first within a language, it would seem. The word “synthetic” means “about the real world.” The question then is simply: Are there any a priori synthetic truths? One may suggest that there are indeed none. The one such that Kant accepted is “Every event has a cause.” This leads us into the meaning of cause and causality, which we shall take up later.

1.3.2 Two Types of Science

The point toward which we are directing the previous discussion is absurdly simple. There are two types of science. The first type is descriptive science, in which man looks at the universe and describes what he sees. This surely characterizes the biology work of Aristotle (384–322 B.C.), the naturalist work of Charles Darwin (1809–1882), all the description of the biological world that we see in a good basic college text on biology, and so on. It is essential to realize that all description is incomplete. We may quote the Existential aphorism: “Essence is the totality of appearances” which we translate to mean: the real nature of an entity that is being observed is given only by “all possible ways of looking at it.” Obviously, we never reach this end and will never do so, because,

even with today's observational techniques and apparatuses, the task is impossible, and also, critically, a highly significant part of science is the development of new ways of looking at things, e.g., via the electron microscope, by the quite remarkable techniques of "chopping up" chromosomes and determining DNA sequences, to give just two, now common, but very new (in the history of science) observation techniques.

The second part of science is the development of theory. This forces us to give a picture (our picture, of course) of the nature of theory in science. Our perception of this, as a concept of theory that was used *before* modern physics and quantum mechanics, is as follows. One observed the world by certain observational procedures. These procedures possessed two critical properties. The first is that the observations made by one trained observer were essentially the same as those made by any other trained observer. If, in fact, two observers do not obtain essentially the same resultant observations, then the actual process of observation used by each has to be examined. If two observers appear to be following the same protocol of measurement and they get different results, then we conclude that the specification of protocol of measurement is incomplete and is susceptible to different implementation by different observers. This is, *of course*, a frequently traveled path of investigation. If a protocol of measurement cannot be specified so that two trained observers cannot obtain essentially the same observation by following the *written* protocol of measurement, then the measurement process is not well-defined and needs further specification. We have used the phrase "essentially the same." We have to include this because much observation consists of placing an observed unit into a category or of attaching a numerical magnitude to the unit being observed. In the former case, it may be that the placing in a category is not entirely reproducible between observers, or even between repeated observations of a unit that is judged on other evidence not to have changed. A simple example of this is observation, say, of a mouse recorded on a film as being normally active, hyperactive or hypoactive; another is classification of individuals who are "mentally ill" as being "organic, psychotic or characterological." Clearly, we are unable to describe all the problems in this area, or even indicate, even superficially what they are, except to give our perception of reports in this area, which is that psychiatric diagnoses are unreliable in terms of agreement of independently acting observers. In giving this, we do not intend to be pejorative: the problems are incredibly difficult, much more so than observing growth of a plant, the endocrinology of an ant, or the behavior of an atom that has been hit by a particular type of particle. In the latter case mentioned, that of attaching a numerical magnitude to an object of observation, it is always the case that there is error of measurement, either of inexplicable variability of result of measuring an object that does not vary (according to all we know), or of measuring to a prespecified degree of "tolerance," as when we say that the height of a human is 69 inches, *meaning* that our judgment is that the height is somewhere between 68.5 and 69.5 inches. Such grouping error of measurement is obviously inevitable; the extent of such error can be diminished by using an improved measurement process but it cannot be eliminated totally. We suggest that this is entirely obvious. If the point is accepted then the implications with respect to the use of continuous probability (or relative frequency) models are clear. Our use of a mathematical distribution such as the Gaussian distribution to represent real observations of a numerical magnitude is an approximation that is convenient for many purposes but misleading for some purposes. Without discussion,

we give our view that one never observes *exactly* a random variable that is Gaussian, and furthermore, that we, finite individuals with finite observational abilities, do not observe events of probability zero, as a *naïve* reading of some statistical theory would suggest. The point here is discussed by Kempthorne and Folks (1971, pp. 258–263) in connection with likelihood ideas.

A second property is that the measurement process itself does not affect the properties being measured, so that the achieved measurement can be regarded solely as a property of the object being measured. In all the common cases of physical measurement this is a property that is assumed, rationally, to be met. In the case of measurement of a mental attribute, it is a property that may be questioned. Certainly, in a psychological or psychiatric interview it cannot be assumed. At the level of measuring properties of elementary particles in physics, there is a fundamental indeterminacy in measuring two attributes; position and momentum, that is formalized in the Heisenberg (1901–1976) uncertainty principle.

1.4 VARIETIES OF THEORY

It is essential to distinguish several types or varieties of theory. Rather than attempt to characterize these by terms, such as weak or strong, which always carry pejorative and derogatory connotations, we shall try to give an idea of what we have in mind.

1.4.1 Two Types of Theory

There are, it seems, two basic types of theory. One type is exemplified by the theories of classical physics. These are dominated by modeling a system of one or more particles through time. One observes attributes, say, a, b, \dots , at times t_1, t_2, \dots . One looks at the resultant data and one surmises that the variables a, b, \dots , are functions of time $a(t), b(t), \dots$. One can then conceptualize that their observations are realizations of functions of time, which we can denote by $A(T), B(T), \dots$, such that these are general relations holding over time, T , which are subject to various mathematical relations, usually involving derivatives and partial derivatives. One then has a formal mathematical problem in the conceptual mathematical variables, which one can solve. Having then obtained an understanding of the mathematical functions $A(T), B(T)$, and so on, one then translates this into functions $a(t), b(t)$, and so on, that are to give predictions of what one will observe with the observable variables. Proof of the validity or rather justification, because there can be no proof, of the process is given by observation, obtaining empirical relationships by data analysis, “translating” these into relationships among the conceptual mathematical variables, deducing the consequences in the mathematical formulation, and checking that these consequences are verified as predictions in the observable world. A general problem underlies the whole of this process, the problem of epistemic correlations; on the one hand, we have the observable real world, with observations given by observation protocols; on the other hand, one has a mathematical theory with mathematical variables; one wishes to use deduction in the mathematical system with formulae for and relationships among the mathematical variables to infer formulae for and relationships among the real-world variables. One

is using what are called epistemic correlations between observable variables of the real world with mathematical variables of the mathematical formalization. This process is so widely used in basic mathematical physics that one uses the same symbols for the real-world variables and the mathematical variables. In other areas of science, one finds mathematical variables used in a mathematical formalization that do not correspond to any real-world variables that can be observed. Under such circumstances, the nature, role, and utility of theory must, surely, be questioned severely.

There is no disagreement that the so-called theories of physics are really and truly theories. We see what may be termed the full mix: observation, data analysis, conceptualization to a mathematically exact theory, developing this theory to mathematical consequences of the elements of the theory, and finally verification by reference to the real observable world.

To exhibit the contrast that we wish to emphasize, we ask the reader to consider some examples: the Aristotelian theory of tragedy, sociological theory, psychological theory, and the theory of plant and animal nutrition. In all these cases, we hold the view that the designation as “theory” is valid. It would seem that it should be quite unnecessary to make this statement, and we would not feel called on to make it if we do not see clear evidence that some individuals educated in the so-called “exact sciences” dismiss what some groups, e.g., sociologists or psychologists, describe as their theories, as being not theories at all but strings of highly imprecise verbal, that is, nonmathematical, “literary” expositions that cannot be given the status of theory.

1.4.2 What Is a Theory?

To form judgment of the question of whether an account of an area of human interest should be accorded the status of theory, it is helpful, we think, to first look at physics. Obviously, we cannot review the progression of theory in any direction, but it is useful, indeed critical, to glance over physical theory. We are told by Russell (1959), that philosophy and science began with Thales of Miletus (624–547 B.C.), who is reported to have said, “All things are made of water,” a theory, even though entirely verbal. Anaximander (610– ca. 546 B.C.) questioned this, “Why choose water?” He said, it appears, that man derives from the fish of the sea—again a theory. For Anaximenes (ca. 570–ca. 500 B.C.), the basic matter was air. Later for Pythagoras (ca. 569–ca. 475 B.C.), the whole of reality could be captured by numbers and mathematics. For Heraclitus (ca. 600 B.C.) the real world consisted of a balanced adjustment of opposing tendencies, then he chose Fire as the primary ingredient. These are mere examples from the succession of theories that were held at one time or another. Somewhat later Leucippus (480–420 B.C.) put forward the theory that the world is made up of “rigid, solid, and indivisible” atoms. This theory was developed by Democritus (460–370 B.C.). Necessarily, we do not enumerate the Socrates—Plato story, which used the theory of ideas, except to give our impression that this was both sterile and highly captivating to the point that it has influenced science strongly over the millennia. Also, rather clearly, its main thrust was towards ethics and the nature of man. Books on the nature of the thoughts of Socrates (469–399 B.C.) and Plato (427–347 B.C.) would easily fill a small library. The scientific ideation of Plato was that everything could be reduced to geometry, which much later was reduced to algebra by Descartes. Next came Aristotle who,

we suggest, was perhaps the first real scientist. He worked on classification of animals and did research in marine biology. The assessment of Aristotle, by the later world, and particularly in comparison to Plato is very mixed—on the one hand, just a pale imitation of Plato and on the other hand the first scientist and philosopher of science, as well as having made vast original contributions to human knowledge. After a dormancy of centuries, apart from some Moslem thinkers, such as Avicenna (980–1037), in opposition to the writings of Aquinas (1225–1274), Roger Bacon (1214–1294) can reasonably be regarded as initiating modern science with the thesis that we must resort to experiment. Bacon was condemned by the Pope and spent 12 years in prison. Again summarizing a huge history, and perhaps rather unreasonably, we see the helio-centric theory of Copernicus (1473–1543), the works of observation of Tycho Brahe (1546–1601), and the data analysis of Kepler (1571–1630), events which must surely be regarded as early and critical in the development of physical science. During the same period Francis Bacon (1560–1626) produced his “*Novum Organum*.” According to Russell (1959) “to replace the evidently bankrupt theory of the syllogism,” Bacon put forward the method of induction. Again “cutting through” a long and fascinating history and talking *only* about science, which was physical science, we see the theories of Boyle (1627–1691), Lavoisier (1743–1794), Faraday (1791–1867), Maxwell (1831–1879), and so on, just to mention a few of the significant names. The point of the present discussion is to indicate the succession of theories—theories that are mathematically based. A curiosity of the present time is that the Einstein (1879–1955) axiom that nothing can exceed in velocity the speed of light is now being questioned, and it seems seriously. So, we see, no axiom of a theory of the real world, no basic proposition about the real world, survives the so-called extreme doubting of Descartes. There is, at base, no single generalization about the real world that should be taken as undoubtable. The life of undoubted generalizations of the past has decreased over the centuries, and much more rapidly so in the twentieth century with relatively huge and growing scientific efforts of mankind.

It is our view, then, that there are varieties of theory. There are systems of the real world that can be idealized into very simple ones, with the aid of ideas such as mass and force. Furthermore, these systems can be isolated from the rest of the world, as in the physics laboratory at the elementary college level and even at a more advanced level such as the now easy experiment of weighting an electron. The same happens in chemistry as is too obvious to need discussion. But the actual world is so much more complicated that to try to place biology, medicine, psychology, and so on in the so-called exact physical science mold is little short of ludicrous.

This rather forthright statement needs, perhaps, substantiation. So we give some obvious examples. Consider plant growth, for instance. We have no problem in being reductionist, that is reducing, in our minds, a plant or a tree, say, to a physical system. A standard college text on plants tells us about the system, roots, stems, leaves, flowers, and so on; we can see, *to some extent*, the vascular system; we can, *in some cases*, feed the plant radioactively tagged chemicals, via the soil in which it grows, and we can follow this material as it progresses through the plant. We know a huge amount about plants, but, *also*, there is a huge amount we do not know; for instance we may ask what “really” goes on in mitochondria, what do the Golgi bodies do? These are merely two examples. Then we know that the growth of plants depends on many types

of nutrients, nitrogen, phosphate, potash, and so-called minor elements; on the amount, nature, and timing of water supply to the plant; on the climate the plant experiences, and so on. Then we have to adjoin the demonstrated experiential *fact* that genetics is important. Perhaps, in the not too distant future, we shall have a technique by which we can determine the whole DNA sequence in the chromosomes of a plant. This will be represented by a string of well-defined symbols, doublets from C, G, A, T. The reductionist hypothesis is, essentially, the hypothesis that if we know everything described above and a huge variety of other aspects not mentioned, then we could explain why one ash tree sheds its leaves three weeks earlier than another ash tree which grows some 70 feet apart. This is plainly silly. We shall never have enough data to establish the types of law that we see in physical science. Even if we had all the data on plants that a group of well-trained biologists regard as relevant, we shall be in the position of trying to model, for simplicity, one attribute such as plant height at maturity in terms of thousands, even millions, of potential explanatory factors.

Now let us take another example—humans. The complexity of the bio-physical system in humans is really quite fantastic. Surely, there is no need to exposit why a pure reductionist attitude and approach cannot be generally followed. We can be reductionist about certain phenomena, such as certain genetic diseases, and many other medical workers could enumerate. But it is plainly silly or ludicrous to attempt to formulate a system of differential equations, say, to explain human growth, these equations, of course, involving all or even a small fraction of the factors that we *know* to be involved.

Obviously, the same sort of discussion can be applied to psychology, sociology, and wildlife studies, to mention just three areas of science.

A consequence of this argument is that in many areas of science the modeling can only be simple, and, often, not even mathematical. As we have said, to use this experiential fact to dismiss many areas of science as not being “real science” is stupid and myopic as well as arrogant. An example is given by Linus Pauling (1901–1994), winner of two Nobel Prizes, one in science and one in peace. The one in science was surely one for reductionist science *in an area that could be reduced*. In his later years, Pauling exposted over the nation his theory that massive dosing of vitamin C will prevent the common cold (Pauling, 1970). Additionally, Pauling has a theory, a *verbal one*, as to why this should happen, this theory relating to the ascent of man in tropical environments (Pauling, 1970). Our point is that Pauling has a theory. One may not like it. One may question it. It is a falsifiable theory, obviously, by means of comparative experimentation.

1.5 THE PROBLEM OF GENERAL SCIENCE

In the pure physical sciences, one can isolate “small” systems from the rest of the world (perhaps at the cost of vast concrete enclosures). Any such small system can be manufactured independently by many scientists. The proof of validity is that different scientists following the same protocol of investigation obtain equivalent results (apart, perhaps, from measurement error).

1.5.1 Two Problems

When we turn to general science, *in contrast*, a first problem is that we cannot manufacture essentially or nearly identical small systems. In human biology, we cannot find two humans who are essentially identical. Even in the case of so-called identical twins, it is the case that the two members will not have experienced the same environment at the same moment of life. What then are we to do to attempt to falsify Pauling's theory, for example. In agriculture, we cannot find two plots of land that are identical. Two plots may look identical to the man in the street, but one can make many physical, biological, and microbiological measurements, to show that the two plots are not identical. Consider, again, a rather advanced and frequently used, surgical procedure, the coronary bypass operation. Can one find two identical humans so that we can have a simple comparative experiment, with one being a control and the other receiving the operation? Obviously, we cannot. Can we model, mathematically, the heart system so that we have a theory to which we can apply the tight deductive approaches of mathematics? Again, obviously not. What then are we to do? Chapters that follow on randomization give *one* suggested process.

A second problem is that we wish to draw conclusions about a population of units, for instance, humans who at present or in the future will have the problems for which coronary bypass surgery is a possible treatment. The standard way (except to Bayesians) to approach examination of a large defined population of units, for instance, humans, is to use the ideas of random sampling—that is, draw a sample at random from the population, examine the sample and attempt to make some sort of inference about the population. But this prescription cannot be applied to the populations of the future for which we wish to “make an inference.” We do not know the set of humans who will be candidates for bypass surgery in the future. How then are we to attempt to form judgments?

1.5.2 The Role of Data Analysis

This problem is, obviously, of vast importance. Unfortunately, it does not seem to be generally recognized to be one in common statistical circles. So we give a little discussion. Given a set of human subjects in 2005, we can perform a comparative experiment. Having performed the comparative experiment, we have to attempt to determine if the response to treatment is in the same direction for all groupings of subjects that we can envisage; i.e., is this so for males and females, for nonsmokers and smokers, for blondes and brunettes, for thin and fat people, and so on. This is *obviously* an impossible prescription to fill. All we can do is to do data analysis in which we look to see if such factors of classification (categorical, ordered categorical, or arithmetically based) give evidence of having explanatory power with respect to outcome of the experiment. We shall discuss this in the later text under the names of “additivity” and concomitant variable analysis. However, we inform the reader that there are no simple answers. The discipline of statistics *suggests* data analysis procedures.

1.5.3 The Problem of Inference

The outcome of such analysis may be, for instance, that thin people do not respond well and that fat people do respond well. Or that individuals of blood group O respond well and individuals of other groups do not. We are merely giving examples. Obviously, in this type of activity “one swallow does not make a spring,” or, more explicitly, one such study is mere evidence, perhaps strongly suggestive. So studies have to be repeated under different naturally occurring circumstances, with different groups of people, for example, South European, North European, African, Oriental, and so on. The outcome, one hopes, is very much the same in all these groups. The wider the groups of people experimented with, the more confidence one will have, in an unquantifiable way, to “extending” a resultant inference to groups not represented in the studies and the more confidence one will have in extending the inference to John Smith in August, 2010, who is 57, white, blood group O, . . . It is clear that the extension of an inference from data to this John Smith is not one that can be made tight. The inference is subjective. It will be made by the controlling physician, and you, the patient, can do nothing except to hope that the physician makes good judgments. One will be able, perhaps, to see data that enables one to quantify formally the judgment ability of the physician; as, for instance, it can be found that he had met 25 cases “like” the one under consideration, made a decision, and then found that he was correct, in some sense, in 24 of the cases. One could formalize this problem somewhat. One could say to the physician, “You, surely, understand coin tossing, so that you understand what a probability of $\frac{1}{16}$ is; simply the probability of getting four heads on four successive tosses of a (tested) penny. So, now please give me your judgment of *your* probability that this proposed operation will benefit me.” We can be quite sure that practicing physicians follow some route of this sort. Presumably and hopefully, this judgment will be based on literature search and on actual experiential facts. Also, however, it is necessarily based on incomplete analogy, expressed by the physician somewhat as follows. “You, John Smith, are a unique individual. No one else has your genetic structure: no one has had your life experiences; no one is in exactly the same configuration as you; but my judgment is that you are sufficiently like the humans in such and such studies, that I feel justified in applying, say, the 95% chance of success observed therein to you.”

1.6 CAUSALITY

It is obviously critical that a general discussion, even if brief and even potentially partially misleading, be given of the idea of causality. This is a topic with history going back to the pre-Socratics; also, we are confident, that it occurs in early non-Greek writings; and it surely occurs in ancient philosophy of the East. We, not only in science, but in nearly all human mental activities, use some concept of causation. As the ensuing discussion indicates the word “cause” has been used for millennia and at present in several senses that are not at all consonant with each other. Because the design of experiments is *directly* aimed at *one* type of causality, it is essential to try to achieve a coherent and *useful* understanding. Underlying every use of the word “cause” is a primeval concept that everything that happens had a cause (cf. the Kantian a priori

synthetic truth).

To begin, let us list some examples, “higgledy-piggledy” of statements using the words “cause, caused, and because of.” Our reason for writing these in haphazard order is a simple one, namely, to exhibit the fact that the concept of cause is used with a recklessness that is really quite appalling. We shall not always complete the statements and shall give . . . to indicate that a proverbial intelligent and educated person, could complete the sentence in various ways.

The sky is blue because . . .

A rainbow is caused by . . .

Such and such happens because of the second law of thermodynamics.

The apple fell because of gravitation.

Genes cause IQ.

Radiation causes cancer.

Socioeconomic status is one of the causal variables of crimes.

The tight money supply causes stagflation.

Bill Jones caused the automobile accident.

The hooter in the factory at Birmingham caused the workers to stop for lunch.

We have day and night because the earth is rotating.

1.6.1 Defining Cause, Causation, and Causality

We now attempt to encapsulate the ideas of Aristotle on cause and causation. We use Runes (1962) as part of our information sources. First on the nature of cause, Aristotle distinguished four interpretations: (1) the material cause out of which something arises; (2) the formal cause, the essence determining the creation of a thing; (3) the efficient cause, a force or agent producing an effect; and (4) the final cause or purpose. This language is surely perplexing. Why should one equate cause and purpose? Obviously, one becomes enmeshed in teleology, that *everything* is in the world for some end, some “telos.” Not unsurprisingly, both the idea that there must be a prime, necessarily single, cause and the idea that life has a “telos,” led Aquinas to one of his proofs for the existence of God, a proof rejected by many outstanding philosophers since, especially Kant. Newton (1643–1727) was a great believer in a concept of causation, that to every effect we must assign a cause, though just what Newton really meant by this obviously innocuous proposition should, we suggest, be considered moot and uncertain.

When we turn to causality, we find ourselves drawn into an even deeper quagmire of words and statements. Causality is the relationship between cause and effect. Given the obvious obscurity about cause and also of effect, it is hard to make progress. We have no doubt, in these days, about accepting the idea that radiation, a “cause,” produces cancer, an effect. M. T. Keeton in the aforementioned *Dictionary of Philosophy* (Runes, 1962) lists nine definitions for causality; they are so important that we have no alternative but to attempt to encapsulate these in very few words:

1. a relation between events, processes or entities in the same time series subject to several conditions;
2. a relationship between events, processes or entities in a time series such that when one occurs the other follows invariably;

3. a relationship, etc., such that one has the efficacy to produce or alter another;
4. a relationship, etc., such that without one, the other could not occur;
5. a relationship between experienced events, processes or entities and extra-experiential events, processes or entities;
6. a relation between a thing and itself (self-causality);
7. a relation between an event, process or entity, and the reason or explanation for it;
8. a relation between an idea and an experience;
9. a principle or category incorporating into experience one of the previous.

If the reader is perplexed with all this, we have great sympathy.

A shorter classification is given by Nowell-Smith (1960). He distinguishes three senses which we attempt to summarize:

- I. Human agency—to cause an event is to perform an action which produces a prechosen outcome;
- II. Causes in Nature—to characterize a natural event that produces a certain prechosen outcome;
- III. Cause as explanation.

This third sense begs the question: “What is explanation?” of course, on which many are curiously silent. This seems to be a conceptualization like I and II without an active agent or a natural agent. It can always be used in answering the question “Why?” It may be a state of affairs, as in the proposition “Height at age 6 causes height at age 12.” The reader may find it useful to place the usages given earlier into categories (1) to (9), or in categories I, II, and III. In spite of the huge use of sense III, the usage is murky. Mill (1806–1873) thought selection of one factor as cause from the whole set of antecedents was arbitrary. Nowell—Smith says: “. . . alternative explanations do not exclude one another; any number of them can be true, and the cause will be relative to the interests and abilities of the investigator.” This, we suggest, “lets the cat out of the bag.” The use of cause in sense III is remarkably vague, fuzzy, and indeterminate.

It is useful, we think, to recall a famous example of Bertrand Russell, the hooters. One is an observer looking at a factory in Birmingham, England; one notices that when a hooter is sounded, the workers stop work for lunch. The event A—“workers stop for lunch” invariably follows the event B—“the hooter sounds at the factory around midday.” We have invariable succession. *Therefore*, event B causes event A. Who can question this? It is surely obvious. But there is, we suggest, a hole. Suppose you are an observer in Glasgow, some 200 miles away. Factory hooters are used there also. Also, you have a screen which you have excellent reason to trust as conveying what is happening at the factory in Birmingham. However, you do not hear sounds at Birmingham. What will you see? The event A*, “the hooter sounds in Glasgow” is invariably followed by the event B, the workers stop for lunch in Birmingham—the

same B as before. Hence it is obvious to you that A^* causes B. Surely! But this is plainly ludicrous. Suppose we really wished to check the proposition that A^* causes B. To do so is easy; most children would, we surmise, tell one what to do: Simply, just arrange for the hooter in Glasgow not to work, e.g., cut off its power or, whatever. Then see if B occurs. What will happen, of course, is that the observer will note that B does occur. The point of this rather silly example is merely to indicate that the idea of inferring causation from invariable succession is a rather hopelessly inadequate and improper act. This problem is, of course, rather simple to look at. Merely stop the hooter in Birmingham and see what happens.

1.6.2 The Role of Comparative Experiments

We do not wish to be derogatory of the various uses of cause and causation. Obviously, humanity has found the various usages useful. The use of cause as explanation as in explaining what happens in a physical process by means of a law—even a conceptual theoretical one rather than a purely inductive generalization—has been fantastically useful. The absence of role of theory has been the perennial and valid criticism of Bacon's "Novum Organum."

This having been said, however, we take the view that cause in sense I, that of human agency, is the critical one for very many, perhaps most, of the concerns of humanity. If we were to radiate humans, would we later find cancer in them? "How can I make the grass on my lawn grow?" is a perennial question of suburban America (and elsewhere, perhaps even in Russia). Our horticulturists have a partial answer: Put on nitrogen. We have experimented, we have done comparative experiments and we found an invariable succession. Event or action, \hat{A} : "Put nitrogen on a lawn" is followed invariably by event \hat{B} : "The grass grows." Outside the sphere of purely theoretical science, the idea underlying this "inference" permeates real world science. The big philosophical movement that underlies this approach is "Pragmatism," formulated by the leading United States epistemologist of all time, C. S. Peirce. It is summarized in the oft-quoted statement of Peirce (1963, p.6):

In order to ascertain the meaning of an intellectual conception one should consider what practical consequences might conceivably result by necessity from the truth of that conception, and the sum of these consequences will constitute the entire meaning of the conception.

We do not claim, really, to understand exactly what Peirce was saying—we merely have glimmerings. (What is the meaning of "by necessity," a phrase so often used in epistemological writings.) We do, however, interpret this and other writings, especially those of John Dewey (1859–1952), to conclude that a necessary and even critical process in all science, whether "pure" or "applied" (an unfortunate but commonly used dichotomy) is the process of comparative experiments. Does vitamin C prevent colds or cause absence of colds? The only way to form a good judgment on this is the controlled experiment in which some individuals receive massive doses of vitamin C and others do not. Then compare the outcomes. It would be plain silly and page-filling to make a list of questions and problems that are attacked by the comparative experiment method. It is hard, even, to think of a human problem of the biophysical nature

and even psychologic, which cannot be attacked by the comparative experiment. We close with just one example. We are all concerned about disease; we all want alleviation of some pain or discomfort at one time or another time. So we have investigators studying methods of intervention—take this pill, have that operation and so on. The last requirement for a submission of a request for licensing a drug use is the clinical trial—a comparative experiment. Without such an experiment that is judged adequate, an application is simply not even considered. The immediate foregoing constitute *the* case for the importance of comparative experiments.

We must now cover a general point. The simple comparative experiment uses experimental units, mice, humans, pieces of steel, etc., and with a proposed treatment, T, includes also a control, C, say, which is nothing but absence of T. One then compares the two groups of outcomes, those following T and those following C. In most *so-called* inexact sciences it is critical that a study possess “adequate controls,” to use a hackneyed but useful expression. It appears necessary to state an “obviousity,” something that is obvious. The behavior under the control may have been established with indefinitely strong validation by previous investigation. If we want to investigate the effect of applied heat to, say, a beaker of sulfuric acid, to take an absurdly simple example, one does not have to do the full comparative experiment of having, say, 6 beakers of acid which merely sit on the bench and 6 beakers to which heat is applied. One knows what will happen to the controls. If we are studying cancer of the colon and contemplate a surgical procedure we do not have to obtain, say, 12 patients and then merely maintain 6, while treating 6 by the surgical procedure. We know, *empirically* what will happen to the controls. In the case of coronary bypass surgery, on the other hand, we do not know how the controls will react, so that we have a huge comparative experiment known as the Coronary Artery Surgery Study (CASS) sponsored by the National Heart, Lung and Blood Institute (see e.g. CASS Principal Investigators, 1983a, b; Rogers et al., 1990; van Belle et al., 2004 (Chapter 20)). We could trace down perhaps 100 such large comparative trials in human medicine, and we could track down thousands in health research organizations, including so-called “drug houses.” We could track down many thousands of comparative experiments in agriculture and biology over the world. And so, on and on. The subject of “Design and Analysis of Experiments” needs no justification by philosophers of science or mathematical statisticians.

Finally, there is a critical point that must be discussed, at least, in a preliminary way. A comparative experiment consists of treating experimental units according to various protocols of experimentation, with necessarily one unit receiving only one protocol. Having done the experiment, what can one conclude? Clearly, nothing more at best than that protocol A led, say, to recovery from such and such a disease; or that the total act of putting 40 pounds of nitrogen per acre on the lawn led to a fine lawn. Having found such a conclusion, it is both necessary and inevitable that one should ask: What in the protocol produced the effect? Recall the hoary, but informative, example that goes as follows: When I drink vodka and tonic, I get drunk; when I drink a scotch and water, I get drunk; when I drink gin and tonic water, I get drunk. What then is the cause of my getting drunk? I ponder the question and come to the conclusion: the only thing common to those interventions that make me drunk is that each intervention includes my drinking water. The example is laughable—but we must interpose—to us with our knowledge. It is, of course, easily questioned by asking: Does drinking water make me

drunk? Obviously, here, we have a historical control with respect to the hypothesis and then, obviously, the causal inference is merely silly.

The example illustrates, however, a remarkably critical aspect. The isolation from a protocol of one particular component as being the “real” causative agent is in some cases a very simple matter, as in our hoary example above. But in general this can be very difficult. The health research field is interesting, and perhaps, exemplar. It is not enough to know that drug X cures an ulcer (supposing that it does). It is regarded as essential to have evidence and hypotheses that are consonant with accepted scientific knowledge of the mode of action of the drug; in what way, does it produce its effect? This brings in, of course, the whole field of pharmacology. This type of epistemological problem pervades science. It is an attempt to justify the efficacy by reference to established scientific laws.

This final point has a rather curious aspect that arises in psychological and sociological experiments. The mere fact of intervention, independently of the nature of the intervention, may produce an effect. This is strongly reminiscent of the possibility that the act of observation alone may produce an effect. The only way of attacking or falsifying such an explanation is, *again*, by comparative experiment appropriately designed.

1.7 THE UPSHOT

We have given a very long discussion of basic ideas. We do not apologize for the length. We have tried, foolishly perhaps, to capture or encapsulate in a few short pages the whole of the intellectual efforts of Mankind to “come to peace with” the unending anxiety of Mankind to understand and control the processes, haphazard though they may seem to be and must have seemed to be to, say, the educated Greek of 1500 B.C. The outcome of this effort is, we suggest, to convince the reader that the role of experiments and in particular that of comparative experiments and of interventional studies are critical in the grand effort. If this is accepted, then there is a clear need for a book on “Design and Analysis of Experiments.”

1.8 WHAT IS AN EXPERIMENT?

To focus briefly, and on a less philosophical basis, on the nature of this book, it is appropriate to ask and discuss this question.

An experiment is “deliberate observation under conditions deliberately arranged by the observer” (Stebbing, 1961, p. 302). This statement is acceptable as a beginning, but it is surely not adequate for general use, because there are many sorts of experiment, just as, for instance, there are many sorts of mammal. It is important to make some sort of classification. If, for instance, John Doe goes to New York to look at Rockefeller Center, this action could be called an experiment, according to the quasi-definition above.

1.8.1 Absolute and Comparative Experiments

One, by now fairly ancient, partition was that into absolute experiments and comparative experiments. Unfortunately, the ideation behind this dichotomy is not at all clear. The determination of the weight of an object, for example, a human, or a sack of sugar is based on the idea that the object has a definite fixed attribute, the weight, which is the result of applying a measuring apparatus to it. The determination of the speed of light, assumed to be a constant attribute of light is obtained by a particular process of measurement. The process of measurement may well be based on theory, which consists of a mathematical structure incorporating mathematical variables that are considered to represent actual real world properties. In any of the examples of this paragraph, the hope is that if the whole process is repeated, one will obtain the same result. If the result is a categorical one such as a color, this may happen. If, however, the result is an arithmetic number, like weight in kg, grams, milligrams, . . . , it is entirely unusual for this to happen. One need only experience the taking of one's weight on a well-graduated balance, e.g., with gradation of $\frac{1}{2}$ or $\frac{1}{4}$ of an ounce. Repetitions of the measurement process will not yield the same number. The assumption that is made, rather uniformly, is that the numbers that are obtained are independent "realizations" of a random variable, which, furthermore, is in fact a standard Gaussian random variable, that is, with zero mean. We place the word realizations in quotes because an actual realization of such a random variable will be an infinite decimal. We say that this assumption has been made by essentially everyone in this area. A few workers have suggested that the appropriate distribution is the double exponential with zero mean.

The idea of repetitions, along with the idea of replication which permeates this area of design and analysis of experiments, is very difficult to characterize. Suppose John Doe makes a measurement at 9:00 a.m. Then has a cup of coffee, and then repeats the measurement at 10:00 a.m. Is this repetition a replication? John Doe at 10: a.m. is different from John Doe at 9:00 a.m. Also, of course, what is being measured may be different over the two times. Curiously, there is very little discussion of the semantic problem that underlies the ideas. Repetition of an observation requires constancy of what is being observed. If what is being observed is not constant, then repetition does not have constant significance. It is clear that a basic component of education in the physical sciences is training in observation so that different observers will obtain "the same result." This does not happen, of course, with any measurement problem of an (assumed) underlying continuous variable. If we assume, for instance, as seems reasonable, that x equals the weight of John Doe at 10:00 a.m. can be any real number, then x cannot be observed. The simplest model of actual observation is that the real line is partitioned by a grid, and actual observation consists of deciding that the value sought lies in a particular cell of the grid. This will not be totally satisfying because the observer will meet cases such as the grid being in intervals of .1 and the observer will meet observations which appear to him to be at a good point. If this happens at all frequently and if, say, the difference between an observation being in $[5, 5.01]$ and being in $[5.01, 5.02]$ is important, then a finer grid must be used.

It is necessary, obviously, that measurements made by scientists agree—that measurements have interpersonal validity. It is not at all obvious that this will happen with a measurement process. So it is necessary that a study be made of the process, by

repeated measurements by the same and different measures. This involves the construction of a design and protocol of such a study.

1.8.2 Three Types of Experiments

We distinguish basically among the following three types of experiments.

Type I: The observation of an assumed constant. Examples are the measurement of

- (i) the velocity of light;
- (ii) the mass of an electron;
- (iii) the gravitational constant;
- (iv) the conductivity of a sample of water.

Any book on chemistry describes a multitude of fixed properties of chemical substances. If a measurement gives variable results in which the variability is greater than that explainable by pure measurement variability, the natural assumption is that the material being measured is not constant.

Type II: The measurement of a property of a population the numbers of which have variability. Obvious examples are

- (i) the average income of the population of families in the USA;
- (ii) the average age of automobiles that can use the roads of the USA;
- (iii) the average of the number of years of education of the adults of the USA;
- (iv) the area in the USA that has been planted to corn in the year 2000. Any book on economics and sociology mentions many such properties.

In Type I, there is strong evidence that there is an underlying constant and the only problem is that there may be, or more generally, will be measurement errors. In Type II, there is the assumption, usually somewhat *well based*, that there is an underlying constant for each member of the population.

The present book is concerned with a very different situation, which we call Type III. It is best exemplified by biological examples, but the same considerations arise throughout all technology, including engineering, and agriculture. Suppose we wish to develop a diet to promote growth in children of, say, age two years. We know from utterly casual observation that children grow at various rates. Suppose we quantify growth by measuring height at two years and at three years of age. We know that we can measure height easily within $\frac{1}{4}$ inch. We know that growth is a very variable process. Some children grow very little from age two to three. Others grow a lot. The variable we are trying to understand and modify is height gain from age two to three. We do not know what diet to use but we have ideas. The only thing we can do is to run an experiment comparing the diets that we judge to merit consideration.

This situation is entirely different from those of Type I and II above. In both of these types there is a true value with the possibility, or, in fact, certainty of measurement or

observation error. In Type III it is impossible to have total replication of any diet. To obtain this, we would have to possess two or more children of age two, who are totally alike and who will be exposed to the same environment from age two to age three. The only thing we can do is to apply each diet to several two year old children and observe them from age two to age three. In doing this, we have replication over children variability. We can obtain measurement variability by repeated measurement using a measurement process that is under statistical control. This Type III experiment is commonly called a comparative experiment because the experimenter is comparing what are naturally called treatments. It is obvious that comparative experiments have been used and, indeed, should be used in all endeavors of critical inquiry, be this in science, and in this we include all sciences, in industry, especially in the manufacturing process, or in government, for developing certain types of social policies (see Northrop, 1948; Scheffler, 1967).

1.9 STATISTICAL INFERENCE

As discussed above and as will become clearer in subsequent chapters, we are concerned about the effects of interventions. The sole purpose of the performance and analysis of experiments is therefore the drawing of inferences about the effects of treatments.

1.9.1 Drawing Inference

We have to present our opinions on what meanings we attach to the term “inference” and to the phrase “drawing of inference.” In *Webster’s Dictionary* (1948, p. 1273) inferences are classified as mediate (= drawn from more than one proposition or premise) or immediate (= drawn from a single premise). The making of an inference is the act of passing from one judgment to another, or from a belief or cognition to a judgment.

The field of statistics has been concerned with drawing judgments from observations.

Let us give a few examples. You, an ordinarily educated citizen, were exposed to various plays, allegedly written by William Shakespeare. This exposure we call the observation. Then you hear that a question has been raised on whether the plays were written by said Shakespeare or by someone else. Your task is to pass from the observation to your judgment.

A second example is that you are to form a judgment on whether the sun will “rise” tomorrow.

A third example is that an observation consists of the result of n tosses of a two-headed coin, which is r heads and $(n - r)$ tails. You are to make a judgment of the result of a $(n + 1)$ th toss.

These three examples have plagued philosophers of science for centuries. They exhibit differences in content. In the first case, the simple judgment is yes or no, though clearly a judgment could be that the author was Bacon, or perhaps others of large extent.

In the second and third cases, a classical answer was to assume that the event was a realization of a binomial trial, p . Then we were to assume that p is a random variable

uniformly distributed over the interval $[0, 1]$. And then to assume that we have observed r successes (heads) in n outcomes. By writing down the joint probability, one can obtain the conditional probability of p , given the observation of r successes from n trials, the posterior distribution of p , as

$$f_{\text{post}}(p)dp \propto p^r(1-p)^{n-r} dp.$$

The posterior expectation of p is then $(r+1)/(n+2)$. This result, that the “probability” of a success after r successes in n trials is $(r+1)/(n+2)$ was known as the law of succession. This “result” received wide support from philosophers and observational scientists for decades, even centuries. The whole story is absurd. What was probability? Why should the “true” probability be “distributed” uniformly over the interval $[0, 1]$? Where did the representation of the result r successes from a trial as a realization of the result of n independent Bernoulli trials come from?

1.9.2 Notions of Probability

Quite a different use of an idea of probability arose in connection with games of chance: for example, with dice tossing and the question of what is *the* probability that a toss of three coins will yield three heads. This question is, of course, totally unanswerable without assuming a probability structure, that is, a class of elementary events with associated probabilities (which were equal). These elementary probabilities were assumed to be frequencies of outcomes. The outcomes were then frequency probabilities in an indefinitely large number of repetitions. The probabilities will not be realized unless the elementary assumed probabilities will be realized in an indefinitely large number of repetitions. This mode of development became a very significant portion of mathematics, particularly in the development of asymptotics. This theory has little bearing on inference except to make a judgment of where the probability model is reasonable.

A very different formulation of probability was developed by J. M. Keynes (1883–1946). He considers our premises to be a set of propositions h , and our conclusion to be a set of propositions a . “If knowledge of h justifies a rational belief in a of degree α , we say that there is a probability-relation of degree α between a and h ” (Keynes, 1921, p. 4). So for Keynes, probability is a degree of rational belief. However, Keynes does not explain what he intends belief to be and not at all what is *rational* belief. He claims that probable beliefs are objective and logical. Keynes then (Chapter IV) discusses a rule by which equiprobability could be established, due to Bernoulli (1654–1705), which he names the Principle of Indifference, according to which if there were several alternatives with no reason for predicating one rather than another, each of the alternatives should have an equal probability. This was discussed by very powerful mathematicians: Borel (1871–1956), Poincaré (1854–1912), and Bertrand (1822–1900). After an unsuccessful attempt to give a forceful presentation of his Principle of Indifference, Keynes (1921, p. 92) says: “The theory of probability, outlined in previous chapters, has serious difficulties to overcome. There is . . . difficulty in measuring or comparing degrees of probability. . . .” He turns to the frequency theory of probability and bases his ideas on those of Venn’s *Logic of Chance* (1962).

Venn (1834–1923) uses as a fundamental concept a series. The variable attributes of a series occur in a certain definite proportion of the whole number of cases in the

series. The probability of an event is the proportion of the event in the series. The origin of the phrase “the frequency theory of probability” is obvious. However, Venn did not discuss how a series should be envisaged. In fact, a series appropriate to a situation can be obtained only by assumption or from a history judged to be relevant and from data analysis.

It is clear that philosophy has not been able to give a well-founded logic of the use of ideas of probability. There have been three developers of ideas of belief calculus of the twentieth century. Harold Jeffreys (1891–1989) gave a set of axioms which included the idea of a prior distribution: let $y = (y_1, y_2, \dots, y_n)$ be the data; assume that this is a realization of a random variable whose probability distribution $p(y|\theta)$ depends on a vector of parameters θ ; then suppose Θ is a random variable with probability distribution $p(\theta)$; then the joint distribution of y and θ is $p(y, \theta) = p(\theta)p(y|\theta)$ which is also equal to $p(\theta|y)p(y)$. Hence $p(\theta|y) = p(y|\theta)p(\theta)/p(y)$. This is the posterior distribution of Θ given y . There are two problems: (1) how do we get $p(y|\theta)$? and (2) how do we get $p(\theta)$? Jeffreys did not address the first. He attempted to obtain a $p(\theta)$ by a logical argument, but failed (naturally). The Jeffreys development is an attempted completion of the Keynes development.

A second development was by F. P. Ramsey (1903–1930), who held the view that probability had to be based on knowledge and could be scaled by reference to frequency obtained by independent tosses of a perfect coin.

A third development was made by L. J. Savage (1917–1971), who used the ideation of Ramsey, in his book, *The Foundations of Statistics* (1954), where he advocated that the prior should be obtained by “introspection.” This work has received great support and has led to the resurgence of “Bayesian inference,” which, incidentally, is a misnaming, because Bayes (1701–1761) obtained his prior by a supplementary experiment.

There have been attempts to justify what are called noninformative prior distributions. Also there have been attempts to relate the choice of prior to the nature of the likelihood functions, $p(y|\theta)$. This function must be obtained from data analysis, so we are back to “square one.”

1.9.3 Variability and Randomization

The need for use of probability ideas arises from the fact that variability of outcome is omnipresent. Furthermore, this variability must be discovered by actual observation of the variability and cannot be discovered by “pure thought.” So, in the beginning must come the obtaining of data and data analysis.

Because we are concerned with design and analysis of experiments we have to consider how we can “live with” variability. We cannot assume that our data are a realization of some convenient stochastic process, e.g., a Gaussian linear model. We shall use randomization and rely on randomization tests of significance and inversions thereof to obtain intervals of uncertainty about effects of treatments. Doing one full test and inversion requires massive computation. However, we find that the randomization distribution of the usual test statistics is closely approximated by the Gaussian linear model distribution of the same statistics. This will be discussed in Chapter 6. The inversions of randomization tests of significance gives then statistical intervals of uncer-

tainty, commonly called (but erroneously) confidence intervals (see, e.g., Kempthorne and Folks, 1971). The point is simply that the probability is ensured essentially by the randomization procedure.

This Page Intentionally Left Blank

CHAPTER 2

Principles of Experimental Design

In subsequent chapters we shall describe in detail various experimental designs, their properties, construction, and analysis. We shall start with very simple designs and then proceed to more complex designs. Each design is based on a certain rationale (and we shall explain the basis for this) and is applicable in certain experimental situations. There are, however, some basic, common principles of experimentation and experimental designs that need to be clearly understood. These principles have to do with the formulation of the problem under investigation, the choice of the experimental design, the execution of the experiment, the analysis of the data, and the interpretation of the results. We shall discuss these principles in general terms in this chapter, leaving the more specific details for later chapters dealing with specific designs.

2.1 CONFIRMATORY AND EXPLORATORY EXPERIMENTS

Most experiments are of an exploratory nature in the following sense. The investigator is interested in finding out what factors have an influence on the outcome of a certain process. For example, one might be interested whether or to what extent the factors concentration of a chemical compound, time of baking, temperature of oven, degree of cooling, and amount of pressure have an effect, either individually and/or jointly, on the breakability of a certain type of cookware. The obvious procedure to follow here is to vary the “levels” of these factors and compare the performance of the various level combinations. How exactly this is to be done is, however, not so obvious. To perform the experiment many decisions have to be made, such as: the choice and number of levels of the various factors, possibly selecting only a subset of all feasible combinations; the choice of the experimental layout as determined partly by the physical conditions, partly by statistical considerations; the choice of measurement of the performance; and the choice of the statistical analysis which is most appropriate for drawing conclusions

for the intended purpose. We shall address these types of questions in later chapters in great detail and discuss the underlying principles so that an investigator can make appropriate decisions for a particular problem at hand.

Experimentation is essentially a sequential process. One experiment leads to another as some insight is gained from a process and new questions are being asked. An exploratory experiment, as described above, may be followed by what we may call a confirmatory experiment. We may, for example, want to compare the “best” procedure found from the exploratory experiment with an established procedure or product and “establish” that the procedure or product is “better” than the old. This in itself is already a well-defined and narrower problem than the one described earlier. As such, it calls for different design considerations. For example, the number of experimental runs may be very important so that the resulting statistical analysis, that is, the statistical test, may achieve a certain desirable power.

To pursue the idea of a confirmatory experiment in a different direction, we may have found the “best” procedure and may want to establish, for process control purposes for example, its statistical properties. We know that process conditions may change and it is important, therefore, to establish the mean performance and the variability associated with the process. For unsatisfactory values this may lead to refinements in the actual production process.

The discussion up to this point has been deliberately vague. It is merely intended to give the reader some idea about kinds of experiments. We urge the reader to think about similar experiments in other fields of investigation and then carry them through the individual steps (to the extent possible) of experimentation which we shall outline in the following sections.

2.2 STEPS OF DESIGNED INVESTIGATIONS

In practical situations many scientific or industrial investigations are doomed to fail. There are many and varied reasons for this, but the most often encountered reason is simply that the investigation was not properly planned. Many investigators fail to understand that careful pre-planning is essential for a successful experiment. This involves going through a number of steps and making decisions at each point before the actual investigation begins.

A schematic presentation depicting the logical steps of scientific and industrial experimentation is given in Figure 2.1.

In the following sections we shall comment on the individual steps and explain their importance in the overall process (for an alternative description of such an approach for industrial experiments see Coleman and Montgomery, 1993). These steps can be divided into two categories: statistical (that is, development of the statistical design, translation into a statistical model, and statistical analysis, which we shall refer to as the “statistical triangle,” indicated by solid lines in Figure 2.1) and nonstatistical in nature. Even though we shall concentrate in this book on the purely statistical aspects of experimentation, it is important to realize that the nonstatistical steps are intimately connected with the statistical steps and require interaction between the subject matter scientist/investigator and the statistician and should not be ignored in any discussion

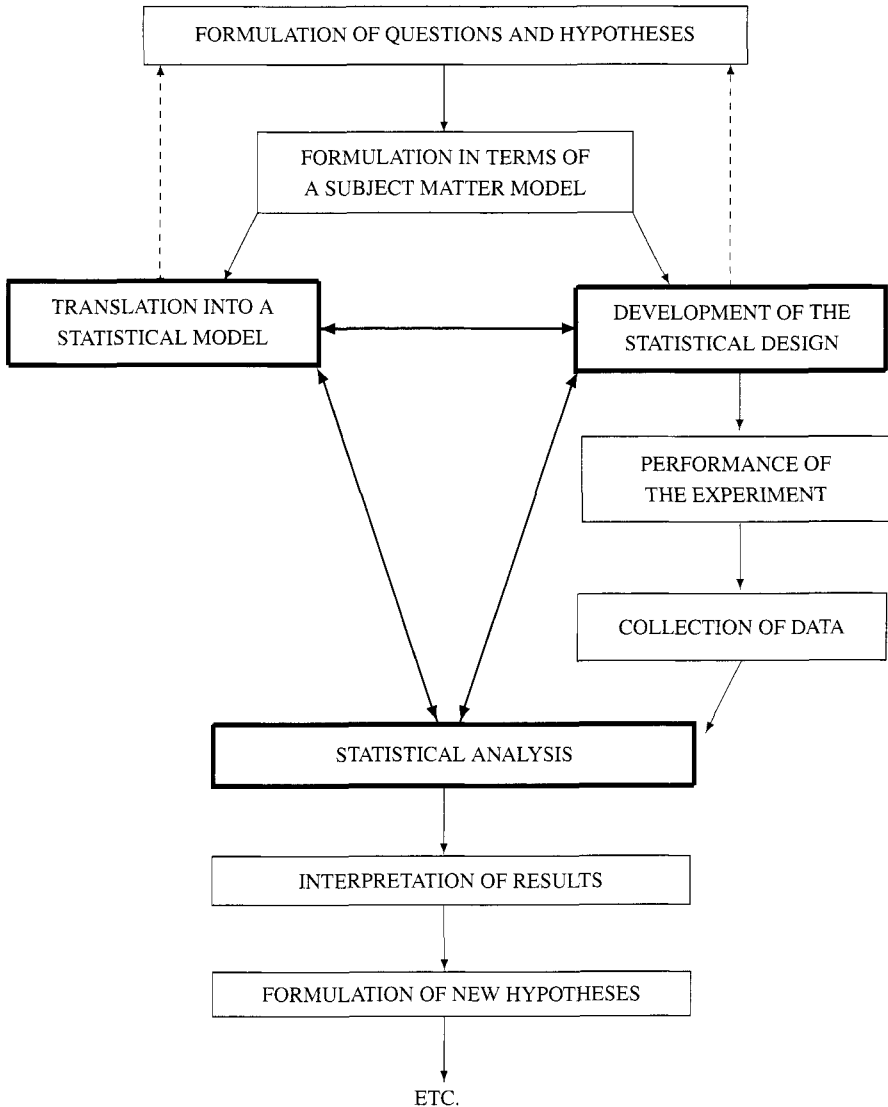


Figure 2.1 Logical steps of scientific experimentation.

of designing an experiment.

2.2.1 Statement of the Problem

Investigation starts often with a simple speculation: “A tree in my garden is hurting; I wonder if it needs more water?” Immediately this leads to questions, such as “How much water should the tree get and how often should it get water?” or “Are there

other deficiencies that need to be corrected and if so, how?" On a more scientific level, speculations and questions of this kind lead to the formulation of a problem: "I wish to determine the best cure for the tree in its present state" or, more in keeping with the topic of this book, "I wish to compare the effectiveness of alternative procedures for curing the tree."

Although this may sound obvious, each scientific investigation must begin with the development and a statement of a problem. This is important not only with respect to subsequent statistical considerations, but also with respect to delimiting the problem to one that can be addressed realistically. All too often experiments are started without a clearly formulated question, purpose or goal in mind, and all too often it is realized too late that such an experiment has been conceived and laid out on too broad a basis leading often to practical difficulties in actually carrying out the experiment. This may lead to a curtailment of the experiment in midstream which may in turn result in an unsatisfactory experiment, that is, one that cannot answer the most important question or questions that the researcher may have. Obviously, other considerations come into play also. These will be outlined in general terms below as a guide to determining reasonable strategies for experimental investigations.

2.2.2 Subject Matter Model

The first step, statement of the problem, is equivalent to the formulation of questions or research and working hypotheses. As mentioned above, such hypotheses must be stated as clearly as possible even though the formulation may not be as precise as that of a statistical hypothesis. Statements such as "I want to compare treatment (procedure) X with treatment (procedure) Y " or "I want to compare several treatments" or "I want to find out which factors have an effect on a certain outcome of a process" are usually quite appropriate. This, in turn, will lead in an obvious way, to the formulation of what we might call a *subject matter model*. By this we simply mean a listing of all the factors that might influence the outcome of the experiment. Such factors will include the treatment factors which are the main objective of the investigation as well as classification (or blocking) factors which are determined by the conditions under which the experiment is performed (see also Section 2.2.4). It cannot be emphasized enough that such a listing is crucial to the whole investigation for the following reasons: (i) It determines, if not completely then to a large extent, the choice of the experimental designs, and (ii) it defines the target population with respect to which inferences can be drawn.

We shall illustrate these ideas with the following examples.

EXAMPLE 2.1: Suppose we are planning a chronic heart failure randomized clinical trial to investigate the effect of carvedilol and metoprolol on the regional vascular responses to adrenergic stimuli (Hryniewicz et al., 2003). In addition to the treatment factor the following factors may need to be considered in deciding on the final trial protocol: gender of subjects; type of subjects, e.g. normal subjects, New York Heart Association class II, III, IV patients; prior or concurrent type of treatment; age of subjects. □

EXAMPLE 2.2: In a study to assess the effect of cognitive behavioral therapy for nocturnal panic (Craske et al., 2005) additional factors to include may be gender, race, marital status, level of education, employment status. □

EXAMPLE 2.3: The importance of mycorrhizal colonization in the establishment and growth of forest trees has long been recognized, and mycorrhizal inoculum is used regularly in replanting (Amaranthus et al., 2005). To further study this effect the following treatment factor may be considered: Type of ectomycorrhizal fungus inoculum, amount of inoculum, type of application. Other classification factors may include: Species of trees, age of seedlings at time of inoculation, environmental conditions in the greenhouse, type and amount of pesticide application. □

EXAMPLE 2.4: A study was conducted to investigate the effects of drinking saline water on farmed deer (Kii and Dryden, 2005). In addition to the treatment factor salinity, the following factors may have to be considered: Deer species, gender of deer, age of deer, location of deer population, other environmental conditions. □

Based upon these factors and possibly others imposed by physical or biological limitations and/or statistical considerations (which will be explained in detail in this book), a suitable experimental design has to be chosen together with an appropriate statistical model. These two steps go hand in hand and once they are established the course of the investigation is pretty well determined and so is the basic statistical analysis. Thus, at this point any reconsideration of the experiment, if needed, should take place (indicated by the broken lines in Figure 2.1). One can think of a number of reasons why such a reevaluation and reformulation of the experiment might become necessary: (i) The experiment, as conceived, has become too big and too complex to be carried out under existing conditions, (ii) the physical limitations imposed by the available experimental material may make it impossible to obtain any or part of the information sought, and (iii) not enough experimental units are available to yield “good” information.

2.2.3 Three Aspects of Design

The point we are trying to make here is that this is the appropriate time to think the experiment through to its logical conclusion before embarking on it. Of crucial importance in this respect is the choice of the experimental design which consists of three components: (i) treatment design, (ii) error-control design, and (iii) sampling and observation design. The *treatment design* determines the treatments to be included in the study: which treatments should we choose and how many? The treatments may be determined by various treatment factors and level combinations of such factors. Then the questions arise how many factors should be used, how many levels should be included for each factor, what is a reasonable range for these levels or what are possible choices for these levels. Not only will this depend on whether the factors are qualitative or quantitative, but also what kind of information is being sought and how that will be reflected in the analysis. It is impossible to give specific guidelines as to how to answer these questions. Each experiment has its own characteristics and demands. General guidelines will, however, become obvious as we discuss in later chapters more specific

designs, both treatment designs and error-control designs.

Aspects of treatment design are closely connected to aspects of error-control design. By *error-control design* we mean the actual arrangement of the treatments in an experimental plan using a rule of assigning the treatments to experimental units, that is, to pieces of experimental material. Examples of such designs are the completely randomized design, randomized (complete or incomplete) block design, Latin square design, etc. (see Chapter 3). The choice of an error-control design depends on the availability of experimental units, the structure of those units, and the precision of estimation desired by the investigator. For example, if the experimental units have a block structure, that is, can be grouped into sets (blocks) of homogeneous units, then some form of randomized block design may be called for. Or if the experimental units exhibit heterogeneity in two directions (as perhaps in a field trial), some form of row-column design (e.g., Latin square, Youden square) may be the most appropriate design. The principle of blocking (see Section 2.5) and the way in which the resulting designs control the error will be explained in later chapters together with reasons for choosing one design over another.

The third component of the experimental design is the *sampling and observation design*. It determines at which level observations are being taken and what kinds of observations are being taken. More precisely it tells us whether the observational units are the same as the experimental units or whether subsampling from the experimental units is to be done. Also, it specifies whether univariate or multivariate observations are to be taken.

As mentioned before and as indicated in Figure 2.1, the development of the experimental design and the formulation of an appropriate statistical model are intimately connected in that the structures of the treatment design, the error-control design, and the sampling and observation design determine essentially the complexity of the statistical model. In the context of this book, the statistical models will be linear models or linearized forms of nonlinear models, more specifically classification and regression models (see Chapter 4). Since this book is concerned mainly with comparative rather than absolute experiments, the linear models used will be classification models incorporating the effects associated with the three component designs discussed above. This will be made clearer as we discuss the various designs.

Having chosen a suitable experimental design, the experiment itself can now be performed. It is worth noting here that although this part of the whole experimental process appears to be nonstatistical in nature, it is crucial that it be carried out in conformance with the statistical requirements for the design chosen. This includes, for example, proper randomization of the treatments to the experimental units and proper replication of the application of the treatments. For this reason it is important for the statistician and the investigator to at least work out and write up a protocol spelling out all the details of the experiment as far as possible. Included in this should also be details about the actual data collection and the measurement process, for example, when data should be collected and what the scales of measurement are.

2.2.4 Modeling the Response

In the actual experiment each treatment factor is represented by different “levels”, that is, different forms or different amounts, such as different types of inoculum, different amounts of inoculum. With regard to the other factors in the subject matter model, the investigator may decide to restrict a factor to just one level, for example, only russa deer, or include several levels, such as patients from different illness severity groups. In the latter case these factors will have to be also included in the ensuing analysis of the data. For this purpose it is important to provide a suitable model of the response data.

To formalize this idea in general terms (which will be made more specific in later chapters) we write

$$\text{Response} = f(\text{Explanatory variables}) + \text{Error}, \quad (2.1)$$

where f represents an unknown function and the explanatory variables refer to treatment and blocking factors as employed in the treatment and error-control designs, respectively.

Among the blocking factors are factors identified by the subject matter model (Section 2.2.2) as essential for defining the target population for purposes of statistical inference. Cox (1984) referred to these factors as *intrinsic* factors. We shall adopt his terminology and divide the blocking factors into intrinsic and *nonspecific* factors, the latter being determined by the necessities of the error-control design, that is, considerations of further reducing heterogeneity of the experimental material. If we denote the set of treatment factors by $\mathcal{X} = \{x_1, x_2, \dots, x_t\}$, the set of intrinsic factors by $\mathcal{Z} = \{z_1, z_2, \dots, z_q\}$, and the set of nonspecific factors by $\mathcal{U} = \{u_1, u_2, \dots, u_s\}$, we can rewrite (2.1) more explicitly as

$$y = f(x_1, x_2, \dots, x_t; z_1, z_2, \dots, z_q; u_1, u_2, \dots, u_s) + e. \quad (2.2)$$

We illustrate the above terminology with the following examples.

EXAMPLE 2.5: We consider an experiment reported by Pearce (1953, 1983) (see also Hinkelmann, 2004) comparing different pruning managements of pear trees. Combinations of different types and amounts of pruning are assigned to individual trees in each of several selected rows of trees. The trees in each row are quite uniform, but there exist row-to-row differences due to environmental conditions. For purposes of inference several varieties of pear trees were included in the experiment. In this setting there are two treatment factors: x_1 = type of pruning and x_2 = amount of pruning, one intrinsic factor: z_1 = variety of pears, and one nonspecific factor: u_1 = rows of trees. The final experimental design is a factorial treatment design (see Section 11.1) in a randomized complete block design (see Sections 9.1 and 9.2). \square

EXAMPLE 2.6: The following description of a clinical trial serves as another illustration. Suppose we want to investigate the effectiveness of different treatments with regard to the elimination of a certain type of skin rash on the human body. The treatment factors are x_1 = concentration of lotion, x_2 = frequency of application. A combination of different levels of each factor defines the medical treatment, and each arm of

each patient included in the trial receives a different treatment. The trial includes male and female patients classified according to disease severity. Thus gender and severity classes represent the intrinsic factors z_1 and z_2 , respectively. The patients represent a nonspecific factor, u_1 . The resulting experimental design consists of a factorial treatment design (see Section 11.1) and some form of incomplete block design (see Section 9.8) as the error-control design. \square

2.2.5 Choosing the Response

In the preceding discussion we have used the term “response” in a generic sense. In many situations it is actually clear what the response or response variable should be. If, for example, we want to determine the effect of different manufacturing processes on the strength of a certain type of plastic tube, then the obvious response is measured in psi, pounds per square inch, needed to destroy the tube. As another example, if we want to assess the effects of different pollutants on a certain crop, it may not be so clear what should be measured. We could measure the growth of the plants at the end of the trial or at the end of the growth period, or the yield of the plants at the end of the growing season, or the amount of damage on the leaves of the plants at the end of the trial.

Our intention here is to point out that the researcher must give careful consideration to the choice of response variable, one that is most meaningful in the context of and most clearly associated with the expected outcome of the experiment. In other words, the response variable should be chosen so that the inferences and results from the experiment can be clearly stated and communicated. In this context, a continuous response variable is preferable to a binary or ordinal variable because it contains more information. On the same grounds, an objective, that is, measurable variable is preferable to a subjective variable. All this depends, of course, on whether we have a choice at all.

There are still other considerations. For example, should we measure the yield of an individual plant or of a bunch of plants? Or, should we measure the pulse rate of an individual at a certain point in time or at several time points within a specified period? Again, this may be determined by the type of inference we want to make concerning the treatments used in the experiment.

2.2.6 Principles of Analysis

Once data have been collected they will be subjected to a statistical analysis in concordance with the experimental design and its associated model. Such analyses will be dealt with in great detail in later chapters. We shall mention at this point only the basic principles involved.

A major aim of analyzing data from designed experiments is to quantify and evaluate the importance of possible sources of variation. This can be achieved through the analysis of variance (ANOVA) associated with the underlying linear model, either in its univariate or multivariate form. The topic of ANOVA will be taken up in great detail in Chapter 4. For purposes of the discussion in this chapter we shall give just a brief outline.

Given observations, y say, from an experiment, the general idea of ANOVA is to partition the “total variability” (or total sum of squares), $SS(\text{Total}) = \Sigma(y - \bar{y})^2$, into component parts as specified by an underlying linear model. Such a model reflects the structure of the observations as determined by the treatment design, the error-control design, and the sampling design [see also (2.3)]. Each design is represented by several sets of effects (parameters). These effects provide a more explicit expression of model (2.2) in the form of so-called main effects of and various interactions among the explanatory variables in (2.2), in addition to one or more error terms (see Section 2.3.2). Suppose there are q such sets altogether. Then, using the method of least squares (see Chapter 4), $SS(\text{Total})$ is partitioned (not necessarily uniquely) as follows

$$SS(\text{Total}) = SS(1) + SS(2) + \cdots + SS(q),$$

where $SS(i)$ represents the sum of squares associated with the i th set of effects ($i = 1, 2, \dots, q$) accounting in some sense for the variation that can be attributed to these effects [see also Section 2.9 for a brief discussion of the partition of the total number of degrees of freedom (d.f.) into the d.f. associated with the individual $SS(i)$]. Of particular interest and importance in our subsequent discussion will be the sums of squares associated with the treatments, and with experimental error.

The ANOVA provides the basic information necessary for making statistical inference either in terms of tests of hypotheses (or tests of significance) or confidence interval estimation. Associated with a sum of squares, $SS(i)$, are the d.f. ν_i , and the mean squares, $MS(i) = SS(i)/\nu_i$. It is the form of the expected mean squares, $E[MS(i)]$, which determines, for example, how tests of hypotheses are performed and how error variances are estimated.

All tests of hypotheses (or significance) and estimation of parametric functions are done in accordance with the aims of the experiment. Thus the statistical results will have to be interpreted in terms of the investigator’s originally formulated hypotheses.

2.3 THE LINEAR MODEL

2.3.1 Three Types of Effects

In our discussion we have pointed out repeatedly the importance of the linear model as it relates to the subject matter model and the experimental design. Such models will be given or derived for all designs presented in later chapters, but it seems useful to give some heuristic arguments here about the form of these models. The general idea is to express the observations, generally denoted by y , in terms of “effects” which contribute to y . These effects or components fall basically into three categories: (i) treatment effects, (ii) design effects, and (iii) error effects.

The treatment effects are a reflection of the intervention procedure or treatment design. The treatment factors listed in \mathcal{X} (see Section 2.2.4) indicate whether a single treatment or combinations of several treatment factors are used. Together with subject matter considerations this will determine which effects, that is main and interaction effects, will be included in the linear model.

The design effects are determined by the explanatory variables included in \mathcal{Z} and \mathcal{U} (see Section 2.2.4). We refer to these effects also as block effects as part of the error-control design.

In addition to pure treatment and design effects we may need to include occasionally treatment \times design interaction effects into the model. These arise from possible interactions between treatment factors in \mathcal{X} and intrinsic factors in \mathcal{Z} (see Section 9.6.8, also Hinkelmann, 2004).

Finally, the error effects, or errors for short, represent different kinds of random variation. Such variation arises in connection with the experimental and observational units (see Section 2.3.2) as well as some aspects of the actual experimentation and data collection. Again, these aspects will be discussed in more specific details in the following chapters (see, in particular, Section 6.3).

We illustrate some of the notions discussed above in the following example.

EXAMPLE 2.7: The objective of a study by Rosen et al. (2005) was to determine whether nitrogen and sulfur fertility affects glucosinolate concentrations in cabbage. The treatment factors were $x_1 = \text{nitrogen} = N$ (at two rates), $x_2 = \text{sulfur} = S$ (at two rates), $x_3 = \text{cultivars}$ (two types: green cabbage and red cabbage). The experiment was set up as a so-called split-plot design (see Chapter 13) with four replications in two years. Thus $z_1 = \text{year}$ as an intrinsic factor and $u_1 = \text{replicate}$ as a nonspecific factor.

The treatment effects included in the model then are: N rate, S rate, N rate \times S rate interaction, cultivar, cultivar \times N rate interaction, cultivar \times S rate interaction, cultivar \times N rate \times S rate interaction. The design effects are year effects and replicate within year effects. In addition, the treatment \times design interaction effects include N rate \times year interaction, S rate \times year interaction, cultivar \times year interaction. \square

2.3.2 Experimental and Observational Units

In order to understand the nature and use of the error effects or error components it is essential to understand the distinction between the (possibly different) units to which treatments are applied and on which observations or measurements are being made. These units are called experimental units and observational or sampling units, respectively. The *experimental unit* (EU) is the piece of experimental material, to which a treatment is assigned and applied. For example, in a clinical trial where different patients are given different drugs, each patient is an EU. If, on the other hand, each patient is given a different ointment on each arm, then each arm constitutes an EU. Associated with an EU is *experimental error*. Such error is a reflection of the fact that EUs are not alike, that is, cannot be replicated exactly. Contributing to experimental error is also our failure to replicate a treatment exactly, that is, instead of administering 15 ppm of a certain substance, as called for in the protocol, we administer to some units 14 ppm or 16 ppm and so on. We refer to this component of the experimental error as *treatment error*.

We emphasize here already that it is very important to always clearly identify the experimental units for a given experimental situation. In the case of several treatment

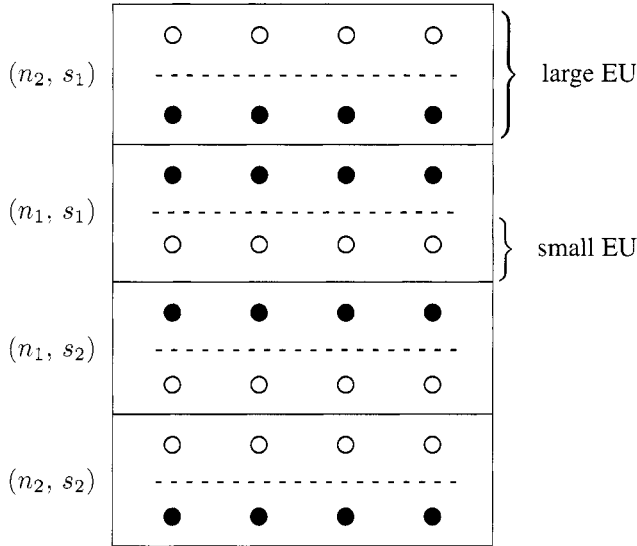


Figure 2.2 Schematic representation of experimental layout.

factors it may happen that different treatment factors are applied independently and separately to different types of experimental units. As a consequence there will then be also different experimental errors, and that becomes important for tests of significance. This is the special characteristic of split-plot type designs (see Chapter 13).

We shall use the experimental setup of the study in Example 2.7 to illustrate the point mentioned above.

EXAMPLE 2.7 (continued): Here we have two types of experimental units. More specifically we have “large” EUs to which combinations of the two rates of N and S are applied, and two “small” EUs within each large EU in each of which one type of cabbage is grown. If we denote the two rates of N by n_1, n_2 , the two rates of S by s_1, s_2 , and a combination of them by (n_i, s_j) ($i, j = 1, 2$), then the essence of the experimental layout for one replicate in one year is illustrated in the schematic representation of Figure 2.2, where the open circles ○ represent green cabbage and the full circles ● represent red cabbage. □

It is important to distinguish between the EU and the observational unit (OU). The *observational* (or *sampling*) *unit* is the unit on which observations, that is, measurements, are made. In many situations EU and OU are identical, but in other situations they are not. For example, in an educational study a class, that is, a collection of students, is the EU as the class as a whole is subjected to a particular teaching method which is the treatment. Observations are made, however, on the individual students in the form of test scores. Then the students are the OUs. Associated with the OU is an *observational* (or *sampling*) *error*, which reflects, among other things, measurement error and also sampling error in that if the experiment were repeated, most likely other students would be part of the study.

To elaborate on this point further we consider again the experiment described in Example 2.7.

EXAMPLE 2.7 (*continued*): One can think of different scenarios of obtaining data in this setting:

- (i) In each row (small EU) we may randomly select one cabbage head and perform the appropriate chemical analysis on it. In this case, at least from a statistical point of view, there is then no distinction between EU and OU. It illustrates, however, the point that there will be sampling error as part of the observational error.
- (ii) We may randomly harvest two cabbage heads per row, chop them and then combine them in a forced air dryer for further analysis. For purposes of statistical analysis the situation is the same as in (i).
- (iii) Two (or more) cabbage heads are randomly harvested from each row and glucosinate extraction is performed on each head. In this case then the EUs and OUs are different: The rows are the EUs and the individual cabbage heads are the OUs. This is an example of what is referred to as *subsampling* (see Sections 3.5 and 6.9). \square

2.3.3 Outline of a Model

In equations (2.1) and (2.2) we have given a very general form of a model for the observations from an experiment. The discussion above suggests that the function f in (2.1) and (2.2) is a linear function. A formal derivation, based on the notion of unit-treatment additivity (see Section 6.3), will be given in later chapters. The idea we want to convey here is that the response after the intervention, which, following an agronomic practice, is often called yield, is made up additively of a unit effect plus a treatment effect plus error effects, e.g., unit or experimental error, observational or measurement and sampling error. The unit effects and experimental error effects are used to model the systematic and random influences, respectively, of the error-control design. Hence we shall refer to the unit effects also as the design effects. They contain the effects of the intrinsic and nonspecific factors of (2.2).

Schematically we write all this as

$$\begin{aligned}
 \text{Observation} &= \text{Design effect} + \text{Treatment effect} \\
 &\quad [+ \text{Design} \times \text{Treatment interaction effect}] \\
 &\quad + \text{Experimental error} \\
 &\quad + \text{Observational error} .
 \end{aligned} \tag{2.3}$$

In (2.3) we have added the design \times treatment interaction effect which may arise in certain experimental situation as a component of interest. We shall elaborate on this model in more detail as we discuss specific examples and different designs – error control designs as well as treatment designs – in later chapters.

2.4 ILLUSTRATING INDIVIDUAL STEPS: STUDY 1

2.4.1 The Questions and Hypotheses

We shall now illustrate the various steps we have described in Section 2.2 and the formulation of a linear model as outlined in Section 2.3 in terms of an example. Suppose an investigator wants to study the effects of air pollutants on seedlings of loblolly pine. The pollutants to be used, singly and in combinations, are O_3 , ozone, and NO_2 , nitrogen dioxide, at levels .00, .05, .10 ppm for 6 hours/day for 28 consecutive days, applied to seedlings at uniform age. The investigator is interested in: (i) comparing the damaging effects of the pollutants and (ii) assessing potential synergistic effects of O_3 and NO_2 . These effects are possibly influenced also by the genetic make-up of the trees, that is, whether they are relatively susceptible or relatively resistant to air pollutants (Kress, Skelly and Hinkelmann, 1982b). More formally, the research hypotheses can then be stated as follows:

- (i) Long-term exposure to O_3 and NO_2 has damaging effects on pine seedlings with respect to growth, mottling, and chlorotic spot symptoms (see e.g., Kress, Skelly and Hinkelmann, 1982a).
- (ii) The amount of damage increases with the level of pollution.
- (iii) A combination of pollutants will exhibit synergistic effects.
- (iv) The amount of damage will also depend on the degree of sensitivity (as determined genetically) of the family from which the seedlings come, one type of family being relatively resistant and one being relatively susceptible.

The observations and measurements will then be determined and influenced by the treatments, that is, the type (combination) of pollutants, the level of pollution, by the genetic background, and for growth by initial height. Other factors such as temperature, amount of light, and humidity, may have to be taken into account depending on the experimental design to be adopted. The seedlings will be exposed to the pollutants in pollution chambers.

2.4.2 The Experiment and a Model

In what follows we shall use language and terminology which is not very precise and intended only to give the reader some feeling for and appreciation of the various concepts we have mentioned so far. More precise formulations will be given in subsequent chapters.

The error-control design depends on the availability and arrangement of the pollution chambers. Suppose the researcher has 18 chambers available to him distributed in a laboratory under uniform environmental conditions, such as heat, light, and humidity. One possible arrangement then would be to randomly assign each treatment, that is, each of the nine possible pollution combinations, to two pollution chambers. In

each chamber a specified number of seedlings, $2s$ say, equally divided between the two families will be exposed in the prescribed way to the assigned pollutant. We shall refer to this as arrangement I.

A linear model associated with this experimental setup might be as follows:

$$y_{ijkl} = \mu + P_i + C_{ij} + F_k + (PF)_{ik} + \varepsilon_{ijk} + \eta_{ijkl}, \quad (2.4)$$

where y_{ijkl} denotes an observation for the l th seedling of the k th type of family in the j th chamber assigned to the i th pollutant, and μ is an overall mean, P_i is the effect of the i th pollutant ($i = 1, 2, \dots, 9$), C_{ij} is the effect of the j th chamber ($j = 1, 2$) assigned to the i th pollutant, F_k is the effect associated with the k th type of family ($k = 1, 2$), $(PF)_{ik}$ is an effect due to the interaction (nonadditivity) between the i th pollutant and the k th type of family, and ε_{ijk} represents an experimental error component and η_{ijkl} represents the observational (or sampling) error ($l = 1, 2, \dots, s$). We note that the experimental error here consists of two components: one component (C_{ij}) arises from the application of the pollutants to different chambers (= EU); the other component (ε_{ijk}) arises in connection with each chamber-family combination as the families, even though they are labeled resistant, say, are not identical, that is, not exactly reproducible as they may be full-sib families produced from different trees. Model (2.4) can be expanded further by making use of the fact that the treatments, that is, pollutants, are actually level combinations of two factors, O_3 and NO_2 , as shown below:

	NO ₂		
O ₃	.00	.05	.10
.00	P_1	P_2	P_3
.05	P_4	P_5	P_6
.10	P_7	P_8	P_9

This structure, which is also referred to as a factorial structure, enables one to partition each effect P_i into O_3 and NO_2 main effects and $O_3 \times NO_2$ interaction. This leads to particular comparisons among the nine pollutants which will enable us to answer, for example, the question posed originally whether O_3 and NO_2 if applied jointly exhibit synergism.

Relating model (2.4) to the terms of our discussion in Section 2.2.4 we note that O_3 and NO_2 represent treatment factors, type of family represents an intrinsic factor, and pollution chamber represents a non-specific factor.

2.4.3 Analysis

An outline of the statistical analysis can be exhibited in an analysis of variance table as given in Table 2.1 (see Chapter 13). This table indicates, again not in very precise terms, which hypotheses can be tested and that these are in agreement with the investigator's aims. More specific hypotheses can be tested using follow-up procedures as described in Chapter 7. Our main point in all of this is that the experiment is designed in such a way that it can provide answers to the questions posed at the outset of the

**Table 2.1 Analysis of Variance for pollution
Arrangement 1**

Source of Variation	d.f.	Research Hypothesis That Can Be Tested
Pollutants	8	Differences among pollutants [see (i)]
O ₃	2	Differences among levels of O ₃ , averaged over NO ₂ [see (ii)]
NO ₂	2	Differences among levels of NO ₂ , averaged over O ₃ [see (ii)]
O ₃ × NO ₂	4	Synergism between O ₃ and NO ₂ [see (iii)]
Chambers (Error 1)	9	
Families	1	
Pollutants × Families	8	Interaction between pollutants and families [see (iv)]
Error 2	9	
Obs. Error	36(<i>s</i> - 1)	
Total	36 <i>s</i> - 1	

experiment. This, of course, does not imply that this is the only way to achieve these objectives. Physical conditions and fiscal considerations may, indeed, dictate another course of action as long as it is consistent with the aims of the experiment. Concerning the performance of the experiment, care must be taken that the treatments, that is, pollutants, are assigned at random to the pollution chambers, and that the seedlings within a chamber are arranged at random or in some rotating fashion for the duration of the experiment. An established protocol controlling other “environmental” conditions will have to be followed. Attention must be given to the evaluation procedures. For example, should foliar symptoms be measured or evaluated on each needle and how, or should each seedling as a whole be rated. How should height growth be measured, from where to where and during which time period?

After the appropriate data have been collected they will be analyzed according to the model outlined above. It is difficult to draw conclusions in the abstract here without any real data, but in light of what has been said before it should be clear how the results from this experiment can be interpreted. The question of synergism can be answered directly. As a result, it is not difficult to imagine that new questions might be raised which then will lead to a new investigation as part of sequential experimentation. The crucial point in designing an experiment is to make sure that the investigator's questions can be answered in the context of the statistical analysis. This means that we must be able to test, in the analysis of variance table, the statistical hypotheses which correspond to the research hypotheses.

2.4.4 Alternative Experimental Setup

To show in terms of the example discussed above how things can go wrong, we consider the following alternative arrangement, referred to as arrangement II. We assign each pollutant combination to two chambers with $2s$ seedlings of one family in one chamber and $2s$ seedlings of the other family in the other chamber. Expressed alternatively, this means that each combination of pollutants and family is randomly assigned to one chamber (this implies that "family" is now a treatment factor). The reader should recognize that this arrangement is, indeed, different from arrangement I. As a consequence, a different linear model will be used to analyze the data. It can be written in the following form:

$$y_{ikl} = \mu + P_i + F_k + (PF)_{ik} + \varepsilon_{ik}^* + \eta_{ikl}, \quad (2.5)$$

where all the terms are as defined before with ε_{ik}^* and η_{ikl} ($l = 1, 2, \dots, 2s$) representing experimental and observational errors. An outline of the associated analysis of variance is given in Table 2.2. The main result here is that there are zero d.f. for experimental error (this follows formally from the position of the total d.f., $36s - 1$, but also from the fact that each treatment combination is assigned to only one chamber (see Section 2.5)) which implies that we cannot test any hypotheses unless we assume that all or parts of the interaction between pollutants and families is negligible. That assumption, however, is not realistic in light of research hypothesis (iv). Hence this arrangement is of no value and should, therefore, not be used. The only reason for mentioning this arrangement then is to emphasize the importance of checking whether a particular arrangement will lead to a statistical model and hence to an analysis which is capable of providing answers to the questions posed by the investigator.

Table 2.2 Analysis of Variance for Pollution Experiment (Arrangement II)

Source of Variation	d.f.
Pollutants	8
O_3	2
NO_2	2
$O_3 \times NO_2$	4
Families	1
Pollutants \times Families	8
$O_3 \times$ Families	2
$NO_2 \times$ Families	2
$O_3 \times NO_2 \times$ Families	4
Expt. Error	0
Obs. Error	$18(2s-1)$
Total	$36s - 1$

2.5 THREE PRINCIPLES OF EXPERIMENTAL DESIGN

In describing the steps of an experiment we have emphasized the statistical aspects, in particular what we have called with reference to Figure 2.1, the “statistical triangle,” namely choice of an experimental design, that is, treatment and error-control design, formulation of an appropriate linear model, and outline of the statistical analysis based on the chosen experimental design and its associated model. To assure validity of the analysis and to increase its sensitivity we have to observe three basic principles which are crucial to any experiment.

The first principle is that of *replication*. By this we mean that each treatment (or some of the treatments) must be applied to several experimental units. In the absence of systematic differences among experimental units treated alike, such replications will enable us to estimate the experimental (random) error against which differences among treatments are judged. (Unreplicated experiments are useful only in certain situations under certain assumptions.)

To ensure validity of the estimate of experimental error we rely on the second principle which is that of *randomization*. It leads to an unbiased estimate of variance as well as an unbiased estimate of treatment differences, that is, estimates that are free from systematic differences due to otherwise uncontrolled variation. We shall comment on the principle of randomization in more detail in Chapter 5 as well as in connection with the individual types of designs. We shall point out then how randomization is to be performed and how that enables one to formulate appropriate linear models and what effect it has on the statistical analysis.

One of the main objectives in choosing an appropriate error-control design is, in fact the reduction of experimental error. In many cases this is achieved by means of the third principle, that of *local control* or *blocking*. The basic idea is to partition the total set of experimental units into subsets (blocks) that are as homogeneous as possible. In this way the effects of nuisance factors which contribute systematic variation to the differences among experimental units can be eliminated. This in turn will lead to a more sensitive analysis since, loosely speaking, the experimental error will be evaluated in each block so generated and then pooled over the whole experiment. Such blocking (by intrinsic and/or nonspecific factors) can occur in various ways and at various stages of the experiment and is dictated by the experimental conditions and the requirements on the desired sensitivity of the experiment. The Latin square design and the split-plot design are examples of more complicated blocking structures which will be discussed in greater detail later in this book. The obvious implication of the present discussion is that the more blocking is being done, the more sensitive the experiment becomes. This is true, however, only up to a certain point and depends on the amount of systematic variability associated with blocking factors, that is to say that it is a function of the given experimental situation and the amount of knowledge one has about it. Also, it should come as no surprise that increased amounts of blocking will invariably lead to more complex experiments, complex from the point of view of execution as well as analysis. All this will become clearer later on.

2.6 THE STATISTICAL TRIANGLE: STUDY 2

In Section 2.2 we have outlined in great detail the various steps essential to designed investigations. These were outlined schematically in Figure 2.1. We have drawn special attention to the intimate relationship between the choice of the experimental design, the associated statistical model, and the resulting statistical analysis. With reference to Figure 2.1, we have called this the “statistical triangle.” Because of its central role in the whole endeavor of scientific experimentation as described in this book, it is important that we give some further discussion along these lines so that the reader can develop an understanding and appreciation of it. We shall do this in terms of a very simple example (Study 2). The idea behind this is to show how, using only heuristic arguments, models for the observations (yields) can be formulated which reflect different experimental situations. The major point we would like to impress upon the reader is that, although all models for the situations described below contain the same components in the form of treatment effect, experimental error and observational error, the roles of the two error components and their associated mean squares in the ANOVA depend heavily and crucially on the underlying experimental plan.

2.6.1 Statement of the Problem

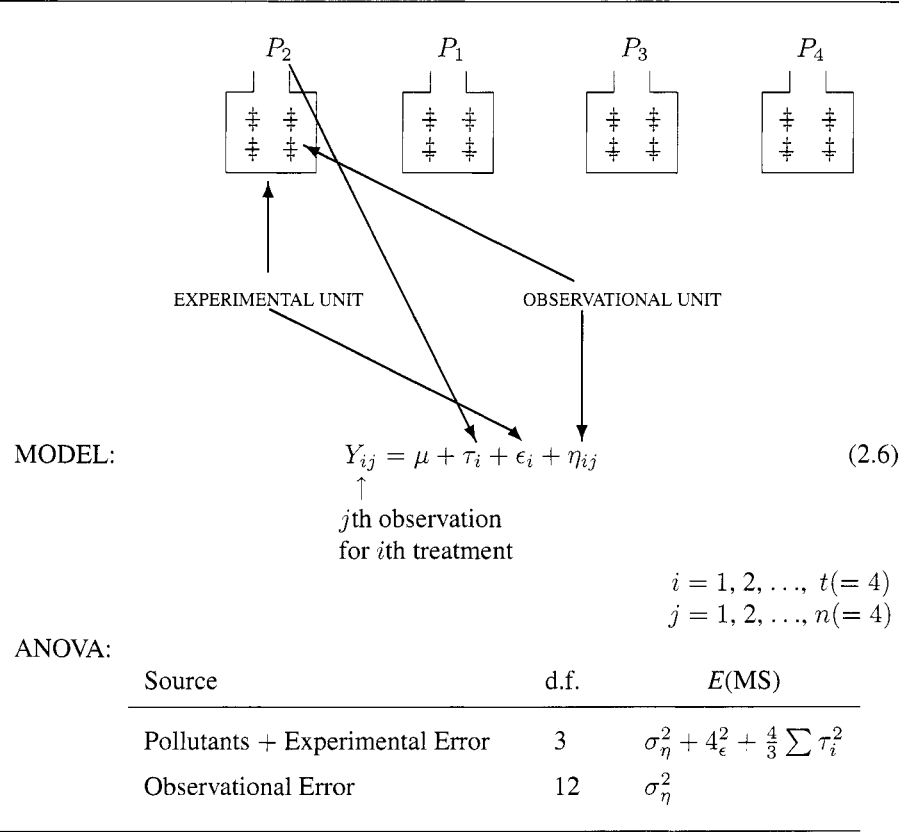
Suppose an investigator wants to study and compare the effects of pollutants on pine seedlings. In addition to charcoal filtered air (P_1) as the control, he includes the following pollutants: ozone (P_2), sulfur dioxide (P_3), and nitrogen dioxide (P_4). This is an exploratory experiment for which he has available four seedlings for each pollutant, that is, 16 seedlings altogether. We shall assume that the seedlings are of the same age and of uniform height, and that a reasonable fumigation protocol has been established and is being carried out properly. The questions we are addressing here are: What are some of the alternative designs for this experiment; what are the corresponding linear models; how can these experiments be analyzed; and most importantly, to what extent can these experiments provide answers to the investigator’s questions?

2.6.2 Four Experimental Situations

In Tables 2.3–2.6 we outline schematically four possible (not necessarily good) experimental plans together with appropriate linear models and an outline of the associated analysis of variance table. There are, obviously, other ways of conducting this experiment, but we shall use the four situations given here to point out differences among them and their associated models and subsequent analyses.

EXPERIMENT I: In experimental situation I (Table 2.3), four pollution chambers are used, each chamber containing four seedlings. The pollutants are randomly assigned to the chambers with four seedlings placed in each chamber. Since a particular pollutant is administered to a chamber, the chamber or, alternatively, the collection of four seedlings constitutes the experimental unit (EU) whereas each individual seedling constitutes the observational (or sampling) unit (OU). As a consequence, the treatment effect and the experimental error are “confounded” with each other, or insepa-

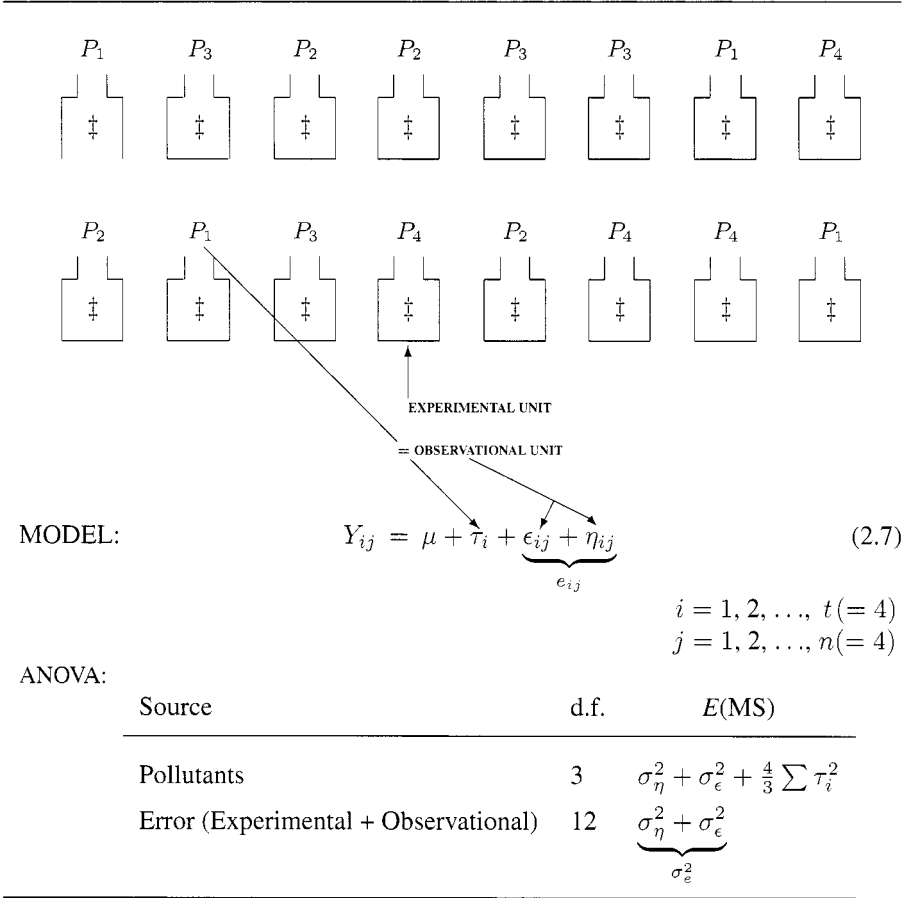
Table 2.3 Study 2: Experimental Situation I



table, which is reflected in the model equation (2.6) in that the treatment effect (τ) and the experimental error (ϵ) have the same subscript. This, of course, leads to the partitioning of the total sum of squares, $SS(\text{Total})$, into only two components, $SS(\text{Pollutants} + \text{Experimental Error})$ and $SS(\text{Observational Error})$. Their expected mean squares, $E(MS)$, which are obtained under assumptions to be discussed in later chapters, make it quite obvious that there is no legitimate error term to test hypotheses about treatment effects, that is, under the null hypothesis that the treatment effects are all identical (and equal to zero), the two MSs do not have the same expected value. From that point of view this experiment is unsatisfactory: It cannot answer the investigator's questions. \square

EXPERIMENT II: In one sense, experimental situation I represents one extreme situation, the other extreme occurring in experimental situation II. Here each seedling is put into a separate pollution chamber, four of which are randomly assigned to each pollutant. Then the EU and OU are identical so that the two associated types of errors cannot be separated from each other as indicated in model equation (2.7). Both errors,

Table 2.4 Study 2: Experimental Situation II

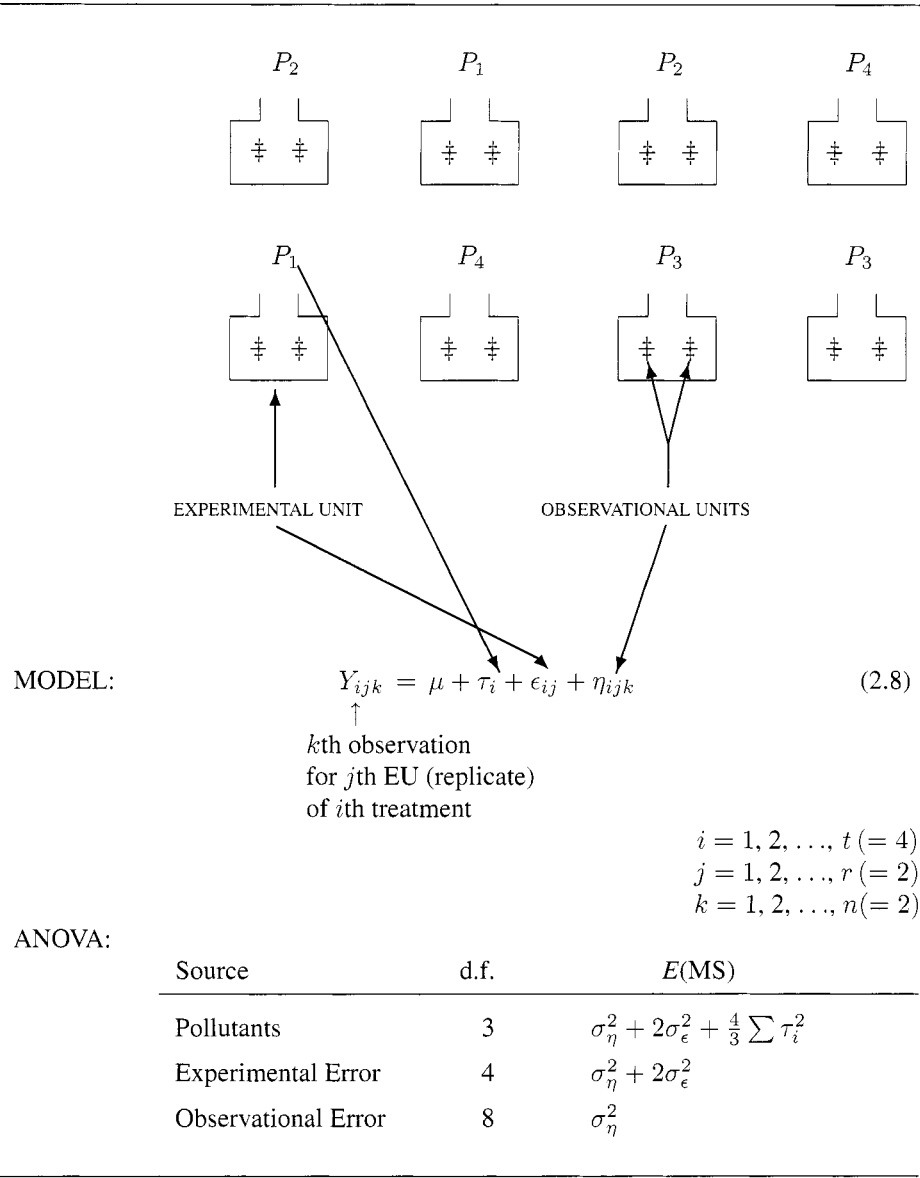


however, can be separated from the treatment effect and hence tests of hypotheses for treatment effects are available (see ANOVA table in Table 2.4). □

EXPERIMENT III: In experimental situation III, two chambers are available for each pollutant so that each chamber contains two seedlings. Variation among chambers (EU) treated with the same pollutant is then a “measure” of experimental error, whereas variation among seedlings (OU) within a chamber is a “measure” of observational or sampling error. Not only are both types of errors separable from each other, but also from the pollutant effects, which is formally expressed in model equation (2.8) as well as in the analysis of variance (see Table 2.5). □

EXPERIMENT IV: Finally, experimental situation IV represents a variation of situation III in that the pollution protocol can be carried out on four pollution chambers with

Table 2.5 Study 2: Experimental Situation III



each pollutant once in the morning (M) and again in the afternoon (A). It is expected that because of the diurnal rhythm of plants, there are systematic differences among the seedlings in the morning and in the afternoon, that is, time of day represents an intrinsic factor. Those systematic differences can be “eliminated” by considering the two sets of four chambers each as blocks. Moreover, this arrangement may lead to a reduction in experimental error (as indicated by σ_{ε}^2* instead of σ_{ε}^2 in Table 2.6). All effects are separable [see model equation (2.9)] and hence this is a suitable experimental procedure. \square

Experimental situations I, II, and III are different versions of a completely randomized design (see Chapter 6) and experimental situation IV represents a randomized complete block design (see Chapter 9). The reader should notice how these different arrangements lead to different models and hence to different analyses. This discussion should also help to bring out the point we have made earlier that it is important to consider the analysis along with the experimental design to ensure that valid statistical inferences can be made. We shall not discuss here which arrangement is best, except to say that arrangement I should not be used, but the use of the other arrangements may be determined entirely by practical considerations and conditions about which we have said nothing here.

2.7 PLANNING THE EXPERIMENT: THINGS TO THINK ABOUT

In the preceding sections we have discussed various aspects of the design process. We have shown how these aspects are interconnected and why it is therefore important to approach the planning of an experiment in a careful and systematic fashion. To emphasize this point we shall summarize below the important features of the individual steps.

1. Statement of the objective (or objectives):

At least a general formulation of the problem to be investigated is essential before proceeding to the next steps. This is even more important if there are multiple objectives which are not to be investigated at the same time.

2. Formulation of the subject-matter model:

The important point here is to prepare a list of all factors that potentially affect the measured response. This involves choosing the treatment factors and identifying possible intrinsic factors. In the interest of keeping the size of the experiment at a reasonable level it may be necessary to restrain some intrinsic factors to just one level, for example female subjects at one age group rather than male and female subjects at different age groups. This will, of course, curtail the applicability of the results concerning the treatment effects, that is, narrow the inference space of the experiment. Another important – and possibly negative – aspect of narrowing the scope of the experiment is the inability of investigating possible interactions between treatment and intrinsic factors.

3. Choosing factor levels:

We are concerned here with both, treatment factors and intrinsic factors. And when we refer to factor “levels” we mean different expressions of that factor, for example, different settings, for instance, 200°C and 300°C, for the factor “temperature”, or different therapies, for instance, radiation and chemotherapy, for the factor “cancer treatment”. For both, treatment and intrinsic factors, it is important to consider carefully how many and which levels we should choose. The choice will affect the type and amount of inferential information as well as the size of the experiment.

For a quantitative treatment factor, for example, it may be important to use more than two levels to assess any possible curvature in the response function (see Section 7.4). Moreover, the levels should be chosen within the practical range for the treatment, including lower and upper limits of the range. Furthermore, the levels should be chosen far enough apart so that a possible difference in response becomes detectable, but not too far apart so that a possible change in response at an intermediate level goes undetected. For example, there may not be any difference between the temperature factor levels 200°C and 225°C, but 200°C and 250°C may be far enough apart for detecting a possible change in response. On the other hand, the levels 200°C and 400°C, may be too far apart because important changes may occur between 250°C and 300°C.

Similar considerations hold also for qualitative factors. For the treatment factors it is usually quite clear which levels to choose, but for the choice of intrinsic factor levels the size of the experiment may become important. For example, in order to investigate the effects of different types of pollutants on plants it may be appropriate to confine oneself to trees or tree seedlings initially. And rather than including in the experiment different species of conifers it may be useful to choose one species from coniferous and one from deciduous trees. A subsequent experiment may then include other plants, such as different types of vegetables.

4. Measuring the response:

The statement of the problem as in 1. above usually not only defines the observational unit (OU) and the response variable but also the way in which the latter should be measured. Such measurements are either continuous or discrete. In some situations different types of measurement are possible, and a decision has to be made which one should be used. For example, to assess the damage due to pollution we may actually measure for each plant the damaged leaf area and the total leaf area and then obtain the percentage of damaged leaf area. This may be rather cumbersome. Alternatively, we may set up a scoring system, say a 5-point score, and then visually assign each plant a score, that is, put it in one of the five categories, that best reflects the amount of damage. Clearly, a continuous measurement is most informative. To approach such a measurement and keep it relatively simple at the same time, we may choose instead of the 5-point system a 10-point system, say, realizing that such a more differentiated subjective scoring

system becomes less repeatable. The other extreme, of course, is a binary system – damaged versus non-damaged – which may be too crude to establish any differences among the pollutants under investigations. In summary, the choice of measurement may be important from a practical as well as from a statistical point of view.

5. Specification of the error-control design:

Identifying intrinsic factors and including several levels of one or more intrinsic factors will already determine to a great extent the type of error-control design that needs to be used, for example, some form of block design (see Chapter 9). To specify the design more explicitly, we need to identify the EUs and OUs and how the treatments are applied to the EUs. This is particularly important if there are several treatment factors. There may be different types of EUs (see Example 2.7) and hence different error-control designs, for example, block design (Chapter 9) versus split-plot design (Chapter 13).

6. Formulating a model and aspects of the analysis:

Although we shall discuss these topics extensively in later chapters, it is important to point out again that mapping out a model and at least parts of the ensuing analysis is a crucial aspect of planning an experiment. These considerations will tell us if and how the research hypothesis can be evaluated within a statistical framework. Among other things we can identify appropriate error terms to test statistical hypotheses or obtain confidence intervals for informative parametric functions. We then can assess whether the error terms are based on a sufficient number of degrees of freedom (d.f.) (see, for example, Sections 6.8 and 6.9.3). In the end these considerations may lead us to conclude that either the experiment as planned is satisfactory or that changes may have to be made to provide for a successful experiment. We cannot emphasize enough that the last point above represents really the culmination of the planning process.

2.8 COOPERATION BETWEEN SCIENTIST AND STATISTICIAN

We have just discussed in detail the various steps of a scientific investigation, with special emphasis on the planning of an experiment. This process requires a close co-operation between the subject-matter, scientist/investigator and the statistician. Below we shall outline some features of such cooperation, paralleling the points discussed in Section 2.7.

1'. Statement of the objective:

Research objectives and hypotheses originate in the context of research activities within a certain subject-matter field. Thus, formulation of such objectives is clearly the primary responsibility of the investigator. It is, however, never too early to contact a statistician if experimental work will be involved. The main

reason for this is to draw attention to the various steps of planning and executing an experiment.

2'. Formulation of the subject-matter model:

This aspect, too, is the primary responsibility of the investigator. This is, however, already a good time for the statistician to raise questions about the desired and possible inference space of the results from the contemplated experiment. Sometimes “dumb” questions by the statistician will help the researcher to clarify and perhaps modify the aims of the experiment. In particular, considerations of potential intrinsic factors will draw attention to the size of the experiment.

3'. Choosing factor levels:

We have argued earlier that in order to make an experiment meaningful it is important to choose the levels of the treatment and intrinsic factors with great care. Here again the statistician has to rely on the subject-matter knowledge of the investigator. It may be desirable from a statistical perspective, for example, to have certain level combinations of the treatment factors present in the experiment, but such combinations may be undesirable or even impossible for biological, physical or chemical reasons, or they may be difficult to achieve for purely practical reasons. In the end, compromises may have to be made to satisfy both statistical and subject-matter considerations without sacrificing the objectives of the experiment.

4'. Measuring the response:

Not many experiments are conducted in a complete vacuum, that is similar experiments have been performed previously. As a consequence, procedures have been agreed upon now to measure the response to treatments. It is generally desirable to conform to such procedures in order to make it possible to make comparisons among the outcomes of different experiments. Precedent and new ways to look upon the results of an experiment may, in fact, lead to using different response measures. This may have to be decided on practical and economical grounds.

5'. Specification of the error-control design:

This aspect of the experiment requires a close collaboration between the investigator and the statistician. Here, questions have to be settled as to how the experiment should actually be performed. At this point a number of questions have to be answered: What are the experimental units (EU)? What are the observational units (OU)? How homogeneous are the EUs? Can and should they be divided into more homogeneous groups (blocks)? Will the experiment be performed at different stages, that is at different times or different places? How will the treatments be assigned to the EUs? Will there be different such assignments? Answers to these and perhaps additional questions will help in selecting one of the error-control designs discussed in later chapters, or help in modifying one of those error-control designs in accordance with the needs of the experiment. An

important aspect of these considerations here is the identification of non-specific factors (see Section 2.2.4) in addition to the already chosen intrinsic factors.

Representing one vertex of the statistical triangle of Figure 2.1, developing the error-control design sets the stage for actually performing the experiment. Therefore, there must be complete agreement between the investigator and the statistician on all points of the design. In order to facilitate communication between both sides and avoid misunderstandings, we strongly recommend to draw a diagram which represents a schematic picture of the physical layout of the experiment, similar to those in Tables 2.3 – 2.6.

6'. Formulating a model and aspects of the analysis:

The model for analyzing the data to be obtained from the experiment is determined in large measure by the treatment and error-control designs, aspects of which we have discussed above. In fact, for each error-control design we shall show in later chapters how a linear model can be derived, and what assumptions have to be made. Such assumptions may involve the nature of certain interactions. Subject-matter knowledge can be of great help in deciding, in particular, which treatment-intrinsic factor interactions may be negligible.

We cannot overemphasize enough how important it is to give careful thought to the basic elements of the statistical analysis and how the various elements relate to the various aspects of the research hypotheses. Hinkelmann (1963) describes an example where from what appeared to be a perfectly logical experimental setup (albeit different from the designs discussed in this book), not all of the researcher's questions could be answered and how the situation could have been rescued if the analysis had been anticipated.

7'. Performing the experiment:

Although the investigator is responsible for performing the experiment, ideally the assisting statistician should be involved, too. Both should make sure that the agreed upon experimental protocol is being followed. For example, it is important to carry out the appropriate treatment randomization in order to avoid bias or confounding. Also, if it turns out that, in spite of careful planning, the experiment cannot be performed in its original form, ways will have to be found to modify or curtail the experiment such that most, if not all, of the original questions can still be answered. An arbitrary curtailment without close consultation between investigator and statistician can lead to undesirable consequences and, indeed, failure of the experiment.

8'. Collecting and recording data:

This is the final step prior to the formal analysis of the data. Not only is it important to collect the data as carefully and completely as possible, following the established protocol, but also to label and organize them in close cooperation in order to facilitate the analysis, typically using some statistical software program. Although so-called missing observations can be handled in many situations by

statistical software packages, this does not provide a license for carelessness, since this may lead to needless complications in the analysis and its interpretation.

2.9 GENERAL PRINCIPLE OF INFERENCE AND TYPES OF STATISTICAL ANALYSES

Our discussion in previous sections has made it clear that the analysis of data obtained from a designed experiment depends very heavily on a linear model which should reflect the structure of the experiment itself. In fact, formulation of such a linear model is a very important aspect in this whole endeavor and we shall return to this problem throughout the book. Once an appropriate linear model has been formulated, the next step will be to obtain the associated analysis of variance or, as the case may be, several analyses of variance. These then provide the basis for making statistical inferences as deemed appropriate for the particular situation at hand.

2.9.1 General Model

As mentioned earlier, a linear model for an experimental design contains generally three types of components: treatment components, design components, and error components. A linear model can be written more formally than (2.2) and 2.3) as follows:

$$\mathbf{Y} = \mathbf{J}\mu + \sum_{i=1}^t \mathbf{X}_i \boldsymbol{\tau}_i + \sum_{j=1}^b \mathbf{U}_j \boldsymbol{\beta}_j + \sum_{k=1}^c \mathbf{Z}_k \boldsymbol{\varepsilon}_k + \sum_{l=1}^d \mathbf{W}_l \boldsymbol{\eta}_l \quad (2.10)$$

where \mathbf{Y} represents an $s \times 1$ vector of observations, μ is an overall mean, and $\boldsymbol{\tau}_i = (\tau_{i1}, \tau_{i2}, \dots, \tau_{ia_i})'$ is an $a_i \times 1$ vector of "treatment effects" ($i = 1, 2, \dots, t$), $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jb_j})'$ is a $b_j \times 1$ vector of "blocking effects" ($j = 1, 2, \dots, b$), $\boldsymbol{\varepsilon}_k = (\varepsilon_{k1}, \varepsilon_{k2}, \dots, \varepsilon_{kc_k})'$ is a $c_k \times 1$ vector of experimental errors ($k = 1, 2, \dots, c$), $\boldsymbol{\eta}_l = (\eta_{l1}, \eta_{l2}, \dots, \eta_{ld_l})'$ is a $d_l \times 1$ vector of observational errors ($l = 1, 2, \dots, d$), \mathbf{J} is an $s \times 1$ vector of unity elements, \mathbf{X}_i , \mathbf{U}_j , \mathbf{Z}_k , \mathbf{W}_l are known matrices of order $s \times a_i$ ($i = 1, 2, \dots, t$), $s \times b_j$ ($j = 1, 2, \dots, b$), $s \times c_k$ ($k = 1, 2, \dots, c$), $s \times d_l$ ($l = 1, 2, \dots, d$), respectively. The matrices \mathbf{X}_i represent the treatment structure, e.g., treatment factors and their interactions (and possibly treatment-intrinsic factor interactions), whereas the matrices \mathbf{U}_j reflect the error-control design aspects, that is, the various blocking devices as suggested by the intrinsic and non-specific factors, and the matrices \mathbf{Z}_k and \mathbf{W}_l reflect the error structure which is partly induced by the blocking devices and various stages of randomization as well as the nature of EUs and OUs and the various types of errors associated with them and with the measurement and observation process.

2.9.2 Outline of the ANOVA

Based upon a model of the form (2.8) we can outline, albeit in not very precise terms, the general structure of the analysis of variance as given in Table 2.7. The basic parti-

tion of $SS(\text{Total})$ is into $SS(\text{Among EUs})$ and $SS(\text{Within EUs})$ with $m - 1$ and $s - m$ d.f., respectively, assuming that there are m EUs. The $SS(\text{Among EUs})$ can then be partitioned further into $SS(\text{Among Treatments})$ with $t. = \sum t_i$ d.f. and $SS(\text{Among EUs Within Treatments})$ with $m - t.$ d.f. Further partitioning of $SS(\text{Among Treatments})$ is possible and sometimes desirable, for example, when one is interested in testing hypotheses about certain treatment contrasts or when the treatments have a factorial structure. Thus, such partitioning is determined largely by the treatment design. The partitioning of the $SS(\text{Among EUs Within Treatments})$ is determined by the error-control design, which leads to various sums of squares associated with blocking factors and associated experimental errors as a function of the different randomizations. The different $SS(\text{Experimental Error})$ will, of course, be used to make statistical inferences about the treatment effects (examples of this will be provided later in the book). Finally, the partitioning of the $SS(\text{Within EUs})$ is determined by the various types of sampling and sub-sampling, that is, by the observational structure.

We shall illustrate this general discussion with Study 1 (Arrangement I) given in Section 2.4.2. Model equation (2.4) can be expressed in the form (2.10) by way of the following correspondences: The fact that with respect to the pollutants the chambers

Model (2.2)	Model (2.8)
P_i	τ_1
F_k	β_1
$(PF)_{ik}$	τ_2
C_{ij}	ε_1
ε_{ijk}	ε_2
η_{ijkl}	η_1

are the EUs implies that $SS(\varepsilon_1)$ provides the appropriate error term (denoted as Error 1 in Table 2.1) for testing hypotheses about pollutant effects. $SS(\varepsilon_2)$, on the other hand, provides the error term (denoted by Error 2 in Table 2.1) for testing hypotheses about pollutant \times family interaction effects. A sampling error is provided by $SS(\eta_1)$.

As illustrated above, one important feature of the analysis of variance is the separation of systematic effects such as treatment and block effects from random or error effects. This is not only important in the context of hypothesis testing but also for establishing confidence intervals and obtaining standard errors for treatment comparisons. Together with model equations of the form (2.10) and properties of (or assumptions about) the error components, the analysis of variance of a properly designed experiment enables us to estimate the variance components $\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2, \dots, \sigma_{\varepsilon_c}^2$ and $\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \dots, \sigma_{\eta_d}^2$ (or linear functions of them) which can then be used as mentioned above. Knowledge about these variance components is quite often useful also to establish further experimental strategies such as determining the appropriate numbers of replications for each treatment and the amount of sampling within EUs. To summarize, statistical inference from experimental data, whether it is in the form of testing or estimation, is based on an underlying linear model. The method of least squares (Chapter 4) is then used to obtain estimates of pertinent parameters as well as the analysis of variance table. In all

Table 2.7 **General Structure of Analysis of Variance for Model (2.10)**

Source		d.f.
Total		$s - 1$
Among EUs		$m - 1$
Among treatments		
Treatment Design	$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_t \end{bmatrix}$	t_1
		t_2
		\vdots
		t_t
Among EUs within treatments		
Error-Control Design	$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_b \\ \epsilon_1 \\ \vdots \\ \epsilon_c \end{bmatrix}$	d_1
		d_2
		\vdots
		d_b
		e_1
		\vdots
		e_c
Within EUs		$s - m$
Observation Design	$\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_d \end{bmatrix}$	o_1
		\vdots
		o_d

this, elements of randomization theory will play an important part.

2.10 OTHER CONSIDERATIONS FOR EXPERIMENTAL DESIGNS

Our main emphasis in this general discussion so far has been that an experimental design, consisting of a treatment design, an error-control design, and an observation (sampling) design, must be chosen in such a way that the investigator’s questions can be answered. In this connection we have elaborated upon the connection between the design and the associated statistical analysis. Now, in many situations a particular scientific objective can be met by different types of experiments. An example of this

was given in Section 2.6. In such a case an obvious question is: Which experiment setup should we choose? Or what is the “best” experimental design for this situation?

Although the question is straightforward, the answer may not always be easy as different criteria have been developed to compare competing designs. We shall mention briefly some of these criteria. Some of them will be discussed in more detail later.

One of the most important criteria is that of optimality or better, variance-optimality. By this we mean maximum precision (in some sense) in estimating linear combinations of treatment effects. Usually a functional of all such variances is minimized which has led to various optimality criteria, such as *A*-optimality, *D*-optimality, or *E*-optimality, (for example, Kiefer, 1959) and it is not always clear which is the best criterion to use (see II. 1.13).

A useful property for a design is that of orthogonality. It allows one to look simply at treatment means for purposes of comparisons. It also leads to a unique analysis of variance the sums of squares of which can be computed easily. All this may not seem very important in these days of high speed electronic computers. It does, however, make the interpretation of results easier and more transparent.

Many of the existing and most commonly used designs are orthogonal, but if orthogonality cannot be achieved, the property of balancedness is often sought. Here we are referring to variance-balanced designs in the sense that normalized treatment comparisons are estimated with the same precision. Other notions of balance exist, particularly in the context of factorial experiments, which are important in the whole discussion of experimental design (for example, Yates, 1935, 1937; Shah, 1958; Preece, 1982; and II.12.5).

There are many other criteria and properties that we could mention here such as connectedness, efficiency, and unbiasedness (see, for example, Federer, 1984), but we shall defer these to later chapters when the need for them will become more apparent.

As pointed out earlier, one of the major objectives in designing an experiment is to estimate comparisons among treatment effects as precisely as possible. This can be achieved in a number of ways such as replication and blocking or refinement of experimental and measurement techniques. One other important device is to use supplementary information in the form of measurements on the OUs which are correlated with the final responses and not affected by the treatments applied to the units. This has the effect of “making the EUs more uniform” and hence of reducing the variability. The statistical technique to be used for this situation is the so-called analysis of covariance (see Chapter 8). It can be used in connection with all types of experimental designs.

Supplementary information may not always be available and if available it may not always be advantageous or it may be too expensive to obtain. It is quite clear that in many instances of designed experiments cost considerations come into the picture. Unfortunately, there is very little of concrete advice that we can offer the reader. Related to this problem we only make one point here: to keep the design as simple as possible as long as it is consonant with the investigator’s objectives. Simplicity is a requirement that affects the execution of the experiment as well as the analysis and the interpretation of results. But simplicity is not an absolute term. The simplest experiment for a given situation may be rather complex indeed.

The notion of simplicity is also tied to another concept which is perhaps the most

important among the ones that we have mentioned: range of validity or target population. If the range of validity is very narrow the experimental design can be rather simple. To give an example, if we want to investigate the effect of ozone at 5 ppm over a period of 30 days at 6 hours/day exposure on loblolly pine seedlings at age 20 days, the target population is rather small and a completely randomized design (similar to the one described in Section 2.6) would be an appropriate design. If, however, we would like to extend the range of validity to natural forests and various forms of pollution, both the treatment design and the error-control design would have to be much more complicated, if indeed it can be done at all without making simplifying assumptions or limiting the target population.

The reader will realize that the preceding discussion has not been exhaustive but rather sketchy. The purpose has been to point out the many facets, statistical and non-statistical, that one must be aware of when designing an experiment. Furthermore, it shows that it is not always possible to meet all the requirements and criteria and that, in fact, some are in conflict with each other so that compromises have to be made. Even though the basic ideas and principles of experimental design are well understood and appreciated, new criteria are being constantly developed (see, for example, Srivastava, 1984) and need be incorporated in this field.

CHAPTER 3

Survey of Experimental Designs and Analyses: A Preview

3.1 INTRODUCTION

In the preceding chapter we have discussed, in general terms, the basic ideas and steps of scientific experimentation, how simple questions and speculations together with knowledge of the subject matter should eventually guide the investigator to a designed experiment which is based on sound statistical principles. We have touched on the major principles of randomization, replication, and blocking and their functions with respect to designing and analyzing an experiment. We shall pursue these ideas in much more detail, of course, as we discuss various forms of statistical designs in subsequent chapters. Our major aim in Volume I is to acquaint the reader, first of all, with a broad variety of error-control designs, treatment designs, and sampling designs so that, given a certain experimental situation, he or she can make a choice among various options, and make that choice intelligently. This means the reader must understand the properties of the various designs, how they can be used to answer the researcher's questions and how a choice of a design from among the possible ones will affect the answer.

The following overview of statistical designs is intended to provide a catalog of designs to be discussed in Volumes I and II and also to describe, albeit somewhat superficially at this point, the hierarchy of error-control designs in terms of their complexity, the nature of treatment designs, the connection between certain types of error-control and treatment designs, and finally, the subtlety of sampling designs. We shall do this in a somewhat schematic way showing the progression from simple to more complex designs. It is useful to keep this in mind as one chooses an appropriate design, be it an error-control or a treatment design, because the choice of a design is often made difficult by conflicting ideas and principles. On the one hand one would like the design to be as simple as possible, mainly for practical reasons, but on the other hand one

would like to account for as many sources of systematic variation as possible, that is, use a more complex design, mainly for reasons of statistical inference. A compromise is often the final result.

We shall conclude this chapter with some remarks about analyzing experimental data.

3.2 ERROR-CONTROL DESIGNS

Table 3.1 gives a list of classes of error-control designs in increasing order of complexity, where complexity is defined by the number of blocking factors for each class. The blocking factors correspond to different sources of systematic variation. The sense then in which the designs “control” the error is through the amount of blocking. Eliminating, that is, blocking for, additional sources of systematic variation (using some intrinsic and/or non-specific factors) will lead to a reduction of the experimental error.

The simplest design is the *completely randomized design* (see Chapter 6) with no blocking factors, that is, assuming essentially homogeneous experimental material (experimental units). Next, a rather large class of designs is that of *randomized block designs* (see Chapter 9). As the name indicates for these designs we have one type of blocking, such as different litters, different breeds, different species, different sources of raw material, different manufacturers, and so on. The specific designs in this class are the complete block design, the generalized block design, and various forms of incomplete block designs which are characterized by the fact that each treatment occurs exactly once in each block, several times in each block, or not in every block, respectively. The complete and generalized block designs have a very simple structure and are easy to analyze and interpret. The incomplete block designs have a more intricate structure. Because of the fact that not every treatment occurs in every block these designs constitute what we shall refer to as *nonorthogonal designs*. Historically that has meant a more complicated analysis (for instance, various forms of analysis of variance) but in today’s computing environment that is no longer true. Nevertheless, these designs which are very versatile and flexible deserve special attention and although they are introduced briefly in Chapter 9, a much more detailed technical discussion of the properties, analysis, and construction is given in Volume II.

The use of intrinsic factors will introduce additional blocking, leading to designs that we shall refer to as *replicated randomized block designs*. As we have already alluded to earlier an important feature of these designs is that they allow the investigation of possible interaction between treatments and some or all intrinsic factors.

An important principle in the design of experiments is that of a Latin square. In its simplest form this is a $t \times t$ row-column array such that every one of t symbols appears exactly once in each row and in each column. In the context of experimental design, the rows and columns refer to two different blocking factors and the t symbols refer to the treatments (see Chapter 10). Thus, compared to the randomized block designs, we have one additional blocking factor. Furthermore, the structure obtained through the Latin square principle is such that the blocking (in two directions) is orthogonal, resulting again in a very simple analysis. The general class of *Latin square type designs* contains several specific designs, such as the *Latin square design* and the *Latin rectan-*

Table 3.1 Hierarchy of Error-Control Designs

Number of Blocking Factors	Class of Designs	Specific Designs
0	Completely Randomized Designs	
1	Randomized Block Designs	Randomized Complete Block Design Generalized Randomized Block Design Incomplete Block Designs: Balanced Incomplete Block Design Balanced Treatment Incomplete Block Design Partially Balanced Incomplete Block Design Lattice Design Extended Block Design Trend-free Block Design
≥ 2	Replicated Randomized Block Designs	
2	Latin Square Type Designs	Latin Square Design Latin Rectangle Incomplete Latin Square Design (Youden Square) Cross-over Design
≥ 3	Replicated Latin Square Designs	
3	Græco-Latin Square Designs	
≥ 3	Mutually Orthogonal Latin Squares	

gle design, which are, except for the degree of blocking, comparable to the complete and generalized block designs, respectively. Corresponding to the incomplete block designs, we now have *incomplete Latin square designs* where, as the name implies, the Latin square principle is not completely satisfied because, for example, the number of columns is less than the number of rows and treatments. This makes the requirement for the design less rigid, but it makes the analysis slightly more complicated. This is a reoccurring theme: We “gain” something on the one hand, but “lose” something on the other hand.

The usefulness of Latin square designs can be enhanced through replications of the basic design. This enables us to take one more blocking factor into account which, of course, widens the inferential basis for the experimental results. Other extensions of the Latin square design using more than two blocking factors lead to designs in the form of *mutually orthogonal Latin squares*. An example of this is the *Græco Latin square design* with three orthogonal blocking factors. Just as for the Latin square design, replications of the basic design will make them more useful.

3.3 TREATMENT DESIGNS

Each of the error-control designs mentioned in the previous section is used to compare t treatments with each other. So far we have not said anything about the nature of the treatments, and it is indeed not necessary to do so. Very often, however, the treatments are chosen to have some structure, in particular a *factorial structure*. This is what we have referred to as the treatment design. Just as the error-control design, the treatment design has to be chosen by the experimenter based upon the goals of the investigation and the experimental material and resources available. The chosen treatment design will then be embedded into an appropriate error-control design.

For factorial treatment structures, we distinguish between *symmetrical (pure) factorial structures* (also referred to as symmetrical (pure) factorial experiments) and *asymmetrical (mixed) factorial structures* or experiments (see Chapter 11). For the symmetrical structure, we have n factors each at s levels, say, where s is an integer. This is also referred to as an s^n factorial. The most useful and practical values for s are 2, 3, and 4. For the asymmetrical factorial structure we have n_1 factors at s_1 levels, n_2 factors at s_2 levels, \dots , n_m factors at s_m levels, where the s_i ($i = 1, 2, \dots, m$) are different integers and $n_i \geq 1$ ($i = 1, 2, \dots, m$). We refer to this as an $s_1^{n_1} \times s_2^{n_2} \times \dots \times s_m^{n_m}$ factorial. An important property of any factorial experiment is that it allows one to study not only the effects of the individual treatment factors, but also the interactions between treatment factors. The usefulness of factorial experiments rests, however, upon the fact that, typically, interactions involving several factors are nonexistent or negligible from a practical point of view.

This observation is of great value in that it allows us often to reduce the size of the experiment by considering only a fraction of all possible treatment (level) combinations. Such an experiment is referred to as a *fractional factorial*. For 2^n , 3^n , and $2^m \times 3^n$ factorials we discuss the basic ideas of a fraction briefly in Chapter 11, but general methods of constructing various types of fractional factorials are discussed extensively in Volume II (see II. 13 – 16). The difficulty in choosing appropriate fractions

is to ensure that essential information about interactions is not being lost.

A schematic overview of these treatment designs and their hierarchy is given in Table 3.2.

3.4 COMBINING IDEAS FROM ERROR-CONTROL AND TREATMENT DESIGNS

It is important to be aware of and understand the properties and purpose of existing error-control and treatment designs in order to make appropriate choices for a particular experiment. But just as these two design aspects are important in and by themselves, it is imperative to understand how error-control designs and treatment designs have influenced each other in generating special error-control designs in particular for factorial experiments. Especially noteworthy here are incomplete block designs for complete factorial or fractional factorial experiments. A brief introduction for 2^n factorial experiments is given in Chapter 11, but the more technical and detailed discussion is deferred to Volume II (see II. 8-12, 13.8). Suffice it to say here only that these designs are constructed by making use of the notion, mentioned above, that certain interactions among treatment factors are negligible and hence information on them can be sacrificed.

Other examples of the interplay between error-control and treatment design are the various forms of *split-plot type designs* (see Chapter 13). This is a large class of designs in which the treatments have a factorial structure, typically with two or three factors. The essential feature of these designs is that the levels of the various factors are applied independently (using independent randomizations) to different types of experimental units by superimposing different error-control designs upon each other. For example, in a simple split-plot design the levels of one factor are applied to “large” experimental units in a randomized complete block design, and the levels of another factor are applied to “smaller” experimental units in a randomized complete block design with the large units representing the blocks, that is, the experimental units for the first factor are split into experimental units for the second factor (hence the name split-plot). Many variations and extensions of this principle exist and are discussed in Chapter 13. As a special case, this contains also so-called *repeated measures designs* (see Chapter 14).

In Chapter 12, we give a brief introduction to *response surface designs*. And even though these designs are not intended for comparative experiments but rather for absolute experiments, the notions of treatment design, that is, factorial experiment, and error-control design, that is, blocking, play a prominent role. As one interesting example of the simultaneous use of error-control and treatment design we mention the so-called Box-Behnken designs, the construction of which is based on essentially superimposing a factorial structure over an incomplete block design.

And finally, we mention a class of designs which are constructed by using the notion of pseudo-factors, that is, by pretending that a factorial structure exists for the treatments when in fact it does not, to construct certain types of incomplete block designs for a large number t of treatments, where t is of the form; $t = k^2$ or $t = k^3$ or $t = k(k-1)$, etc., for some integer k . These are referred to as *Lattice designs* and are of practical value for agronomic experiments (see II. 18).

Table 3.2 Hierarchy of Treatment Designs

Type of Factorial	Number of Factors	Number of Levels	Number of Treatments
Symmetrical (or pure)	n	s (s = prime or prime power)	s^n
Asymmetrical (or mixed)	$n_1 + n_2 + \dots + n_m$	s_1, s_2, \dots, s_m	$s_1^{n_1} \cdot s_2^{n_2} \cdot \dots \cdot s_m^{n_m}$
Fractional (symmetrical)	n	s	$s^{n-\ell}$
Fractional (asymmetrical)	$n_1 + n_2 + \dots + n_m$	s_1, s_2, \dots, s_m	$s_1^{n_1-\ell_1} \cdot s_2^{n_2-\ell_2} \cdot \dots \cdot s_m^{n_m-\ell_m}$

Table 3.3 Hierarchy of Sampling Designs

Type of Design	No. of EUs per Treatment	No. of Subsamples per EU	No. of Sub-samples per Subsample	No. of Observations per Treatment
No subsampling (i.e., EU = OU)	r	1	—	r
Subsampling	r'	n	—	$r'n$
Sub-subsampling	r'	n'	m	$r'n'm$

3.5 SAMPLING DESIGNS

In Section 2.3.2, we have discussed the importance of the notions of experimental units, EU, and observational (or sampling) units, OU. In many experimental situations, the EU and OU are identical. However, if they are not the same, then this needs to be recognized and reflected in the analysis (an example of this is given in Section 2.6.2). Such a situation is referred to as *subsampling* and it can occur in connection with any error-control design. We shall discuss the consequences of subsampling in detail for the completely randomized design (see Section 6.9). The same arguments apply then to all other error-control designs discussed in this book.

The most important feature of an error-control design with subsampling is that it allows the separation of experimental error and observational (or sampling) error, or more precisely, the separation, that is, separate estimation of the experimental error variance and the observational error variance. We shall discuss the statistical implications of this fact in connection with making inference about treatment effects and comparisons among treatment effects. The possibility of being able to estimate the two types of variances may prove to be useful to the investigator in assessing the “quality” of the experimental and observational (measurement) procedures. Large variances may lead to a closer look at and, hence, to possible refinements of one or the other or both procedures.

The notion of subsampling can obviously be extended to more than one level, for example, for each experimental unit we may have several sampling units and then for each sampling unit we may have several observational units. We refer to this situation as *sub-subsampling* for obvious reasons. As an example, consider an individual as the experimental unit receiving a particular treatment; several blood samples, constituting the sampling units, are taken at one time from this individual, and duplicate determinations of, say, the blood sugar level are made, each determination representing an observational unit. Such a scheme would enable the investigator to assess the variability due to three sources: experimental, sampling, and observational.

Theoretically this can be extended even further, but this does not provide any new insight from the point of view of experimental design as discussed in this book.

A schematic representation of the sampling designs described above is given in Table 3.3.

3.6 ANALYSIS AND STATISTICAL SOFTWARE

Following chapters will show that the notion of randomization is not only fundamental to physically performing the experiment but also to analyzing the data from such an experiment. This is accomplished by introducing the notion of *design random variables* which are then used to obtain a *derived linear model* which reflects the randomization procedure used for a particular error-control design. Such a model and the ensuing analysis of variance will then be used to formulate the *randomization analysis* due to R. A. Fisher (1926, 1935). This is a nonparametric analysis and, hence, does not depend on the often quoted normality assumption for experimental data.

Although we advocate the randomization analysis as the proper analysis, it is often met with practical difficulties, because for most situations the number of possible randomizations becomes extremely large. We shall show how the randomization analysis, that is, randomization tests, can be approximated by appropriate F -tests. At this point we shall then make use of existing statistical software for purposes of analysis, always keeping in mind, however, that this represents an approximation only, albeit a "good" approximation. Among the available statistical software packages we have chosen SAS[®], a Statistical Analysis System (SAS Institute, Inc., 2002–2003), the use of which will be illustrated through numerous examples.

3.7 SUMMARY

The preceding discussion and enumeration of classes of error-control and treatment designs and combinations thereof by no means exhaust the list of available designs. Many speciality designs have been constructed and it would be impossible to list and discuss them all. We have, however, mentioned and we shall discuss in subsequent chapters the major classes of designs and their properties, how they are constructed, how they are analyzed, and how they are applied. The major point here is that for many experimental situations special designs may have to be constructed and that this can be done easily by having a firm understanding of the notions of blocking, incomplete blocks, the Latin square principle, the split-unit (split-plot) principle, and factorial treatment structure. They are the building blocks of almost all designs, and it is the aim of this book to elucidate them in a rigorous way, emphasizing the mathematical and statistical aspects.

This Page Intentionally Left Blank

CHAPTER 4

Linear Model Theory

4.1 INTRODUCTION

4.1.1 The Concept of a Model

The greatest intellectual achievement, perhaps, of the twentieth century has been the development of the concept of model and the use of that concept. A model is an explanation of observables in terms of observables. Explanations are of various types. The most simple is, surely, the notion of descriptive explanation; so, for instance, to use an ancient statistical example, height and weight of human adults of one or the other sex follow approximately a bivariate normal distribution. Or at an even more elementary level, with s denoting distance and t being time then with certain units of measurement, $s = \frac{1}{2}gt^2$, where g is a constant of gravitivity. Models can be static in the sense that they describe a situation. They can be dynamic in the sense that they tell us, given the truth of the model, what will happen in, say, the future, the prime example being those arising from differential equations, such as $dy/dt = at + b$, where t indicates time. The whole area of differential equations and partial differential equations is concerned with what may be reasonably called dynamic models so that from a starting point, the differential equation, one can, hopefully, obtain the solution which tells us the outcome over the relevant space, for instance, physical space and time. Models can be classified in another way as being merely explanatory or causal. If we envisage a variable y as being affected by a variable x , and we can, furthermore, envisage a comparative experiment in which the variable x is controllable and is observed at various prechosen levels, then we can reasonably regard a resulting explanation, $y = f(x)$, where $f(x)$, as a special case of (2.2), is some function such as ax , $\ln x$, $\sin x$, or whatever, as being a causal explanation or a causal model. Clearly, the imputation of the explanation or model having a causal basis must in the last resort have an experimental, that is, interventional, basis.

An approximate model is one in which a variable (of arbitrarily general form) is approximated by a function of variables deemed appropriate, or merely being available. So, for instance, we may have the variable y denoting the weight of a child, and we

have a mathematical formula

$$y = \beta_0 + \beta_1 x + \beta_2 z$$

in which we approximately describe y in terms of two variables, $x = \text{sex}$, and $z = \text{age}$. As we shall see later, a very common situation is that we have a vector variable \mathbf{y} , say $n \times 1$, which we wish, for one reason or another to describe approximately in terms of other vector variables, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$, and we seek a description

$$\mathbf{y} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_p \mathbf{x}_p.$$

In contrast to the notion of the previous paragraph which is solely in the realm of approximation and approximation theory, we have the notion of a stochastic model. In this case we have, for example, a random variable, Y , which has a probability distribution, and we wish to describe the distribution of Y ; for instance, that Y follows $N(\mu, \sigma^2)$, the normal distribution with mean μ and variance σ^2 . There is, of course, no limit to the nature and complexity of such stochastic models. One that is surely easy to imagine and think about is that we have a sequence of random variables Y_1, Y_2, \dots and we wish to characterize in some way the joint distribution of the random variables. In such problems, we may have some variables that are considered, in our proposed explanation, to be predetermined, or exogenous, the variables for which we need an explanation being called endogenous. In the case of stochastic models general development involves two entirely different directions, with different mathematical techniques. It may be that we have a basic probability model that is specified mathematically, and we then have to derive by mathematical reasoning the probability distribution of variables that result from numerical processes from the variables in the basic and given probability model. All this leads, of course, and as many readers will realize, to the whole realm of probability theory, stochastic processes, and so on. This is one of the two directions and it requires certain easy and many not-so-easy types of mathematical analysis and technique. The other direction is that we have to envisage, by one route or another (perhaps rather naive and perhaps very sophisticated mathematically) a probability distribution, perhaps specified only to a partial extent, or perhaps completely. Then our task is to do one or both of the following: (a) Obtain observations according to some investigative plan (or even with no plan at all, except that we accept what our observation process gives) and then (b) follow procedures of statistical analysis to estimate, that is, form judgments, preferably, in objective ways, of aspects of the probability distributions, with the assumptions that our given data comprise realizations of random variables.

What we have described above, characterizes, in a sense, all that goes on in all branches of science, though what we have given is a short picture that would require huge amount of writing to exposit in reasonable detail.

What is really happening in this whole broadly specified process is a two-branched operation. In one branch, one abstracts what one envisages as being relevant variables into mathematical entities, variables of one sort or another, one abstracts properties of these real-world variables, and one develops consequences of this mathematical structure that one has abstracted. Then one examines the real world, and one has measurement processes: one decides, or merely hopes, that the result of a measurement

process on the real world, say X , can be regarded as a correlate of an element x in the associated mathematical abstracted structure. In other words, one makes epistemic correlations of real world observables to entities in the associated mathematical structure. In general, the whole process is very complicated. Just to take a very simple case, consider temperature, as a measurable attribute of a specimen, and temperature as it appears in some mathematically formulated theory. The process of developing modes of observation and drawing on measurement protocols is itself a result, nearly always, of some interplay of formal (or perhaps, very informal) theory and observation.

4.1.2 Comparative and Absolute Experiments

We are concerned with the area of design and analysis of experiments, and more specifically, with comparative experiments. The use of the restricting adjective “comparative” is very natural in that we are concerned with entities called treatments which can be applied to experimental units, for example, children, cows, plots of lawn, and pieces of steel, and we are concerned with determining differences between treatments with respect to outcome or response variables, which we shall, often, call yields; for example, with children under different treatments beginning at the age of 6, we are interested in height and weight at age 8, the latter being yields or outcomes or response variables.

In contrast to the comparative experiment we have the absolute experiment. In this case we have observations on a presumed constant (or set of constants) and our task is to determine it; for example, the charge on an electron, or the life curve of a species or race of mice.

The most widely used approach to the comparative experiment, and to a large extent, to the absolute experiment, is the use of linear models to which we now turn.

4.2 REPRESENTATION OF LINEAR MODELS

We suppose that we have a variable y to be explained in terms of variables x_1, x_2, \dots, x_p . We have units or entities, such as, human subjects, plots of land, and pieces of steel, on which we have observed each of the variables y, x_1, x_2, \dots, x_p . Suppose we have observed n units. We may then represent our data set as an $n \times 1$ vector \mathbf{y} , $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ which is to be explained by means of the columns of what we call the *model matrix*

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \cdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p),$$

where $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ are vectors. A linear model is given by the equation

$$\mathbf{y} \doteq \mathbf{X}\boldsymbol{\beta} = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \cdots + \beta_p\mathbf{x}_p, \quad (4.1)$$

in which the coefficients $\beta_1, \beta_2, \dots, \beta_p$ are either given or are to be determined, and there are no relationships among the coefficients $\beta_1, \beta_2, \dots, \beta_p$. We use \doteq to be a

shorthand for “approximately described by.” In contrast to a linear model, we might have occasion, for instance, to consider the model

$$\mathbf{y} \doteq \beta_1 \mathbf{x}_1 + \beta_1^2 \mathbf{x}_2 + \beta_3 \mathbf{x}_3,$$

which is nonlinear in the parameters β_1 and β_3 . Equation 4.1 will be used to encompass two types of model:

- (a) *An approximative model* in which we wish to represent a given vector \mathbf{y} by a linear form in $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$, so that the problem is then strictly one of defining a distance between two $n \times 1$ vectors, say \mathbf{y} and \mathbf{z} and then obtaining that \mathbf{z} which is a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ which is nearest to \mathbf{y} , a problem that clearly lies in approximation theory, perhaps elementary.
- (b) *A stochastic linear model* in which \mathbf{y} is a random vector and

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

where $\mathbf{X}\beta$ is some fixed vector, to be estimated in one way or another and \mathbf{e} is a random vector following some distribution.

4.3 FUNCTIONAL AND CLASSIFICATORY LINEAR MODELS

4.3.1 Functional Models

If we measure, say humans of age 21, with y being weight, x_1 height, x_2 being adult height of the male parent and if we wish to explain or fit y by means of a model on x_1 and x_2 , the values of the explanatory variables can take a continuum of values. It is useful to give such models the name *functional models*.

A general problem that arises in observational studies in which one merely observes the explanatory variables and these are continuous variables is when the explanatory vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ are nearly linearly related by one or more relations of the sort

$$\gamma_1 \mathbf{x}_1 + \gamma_2 \mathbf{x}_2 + \dots + \gamma_p \mathbf{x}_p = \mathbf{0},$$

where $\gamma_1, \gamma_2, \dots, \gamma_p$ are constants and $\mathbf{0}$ is the $n \times 1$ vector of zeros. This is called the problem of *multicollinearity*. It leads to considerable difficulties (for example, Myers, 1990). This problem is of great concern when the explanatory variables are numerous and related in some way not necessarily known or even partially understood as in many economic studies, for example.

4.3.2 Classificatory Models

In contrast to the previous case, the individuals with y -values which we wish to fit by the model may be classified according to factors of classification: for example, with an experiment in blocks and treatments (see Chapter 9) the observational units may be

classified by the block classificatory factor and the treatment classificatory factor. In such cases a *classificatory linear model* is considered of the type

$$y \doteq \text{block effect} + \text{treatment effect}$$

or more conveniently for many purposes

$$y \doteq \mu + \text{block effect} + \text{treatment effect},$$

where μ is some (appropriately defined) constant.

Formal representation of this as a linear model is achieved by the following type of language. Let the units be indexed by $u = 1, 2, \dots, n$. Let $x(u, i) = 1$ if unit u is in block i and let $z(u, j) = 1$ if unit u has treatment j , these variables being otherwise equal to zero. Then a linear classificatory model in simple scalar form is

$$\begin{aligned} y_u &\doteq x(u, 1)\beta_1 + x(u, 2)\beta_2 + \dots + x(u, b)\beta_b \\ &\quad + z(u, 1)\tau_1 + z(u, 2)\tau_2 + \dots + z(u, t)\tau_t \\ &= \sum_{i=1}^b x(u, i)\beta_i + \sum_{j=1}^t z(u, j)\tau_j, \end{aligned}$$

where β_i ($i = 1, 2, \dots, b$) represents the effect of the i th block and τ_j ($j = 1, 2, \dots, t$) the effect of the j th treatment. In matrix form this can be written as

$$\mathbf{y} \doteq \mathbf{X}_\beta \boldsymbol{\beta} + \mathbf{X}_\tau \boldsymbol{\tau},$$

where \mathbf{X}_β and \mathbf{X}_τ are $n \times b$ and $n \times t$ matrices, respectively, of known constants and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_b)'$, $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_t)'$ are vectors of unknown parameters. One may wish to include a constant term and would then have

$$\mathbf{y} \doteq \mathbf{J}_n \mu + \mathbf{X}_\beta \boldsymbol{\beta} + \mathbf{X}_\tau \boldsymbol{\tau}, \quad (4.2)$$

where \mathbf{J}_n is the $n \times 1$ vector of unities.

The significant aspect of models of the form (4.2) is that every element of \mathbf{X} , the model matrix, is equal to 0 or 1. Then, additionally, because every unit is in one and only one of the blocks and receives one and only one treatment, we have relationships like

$$\mathbf{X}_\beta \mathbf{J}_b = \mathbf{J}_n, \quad \mathbf{X}_\tau \mathbf{J}_t = \mathbf{J}_n.$$

From one point of view we have in the model multicollinearity of a very simple type resulting from the fact that the model contains contributions, combining additively, from subsets of the data resulting from imposition of classifications into disjoint subsets.

4.3.3 Models with Classificatory and Functional Components

A more general class of linear models has both functional and classificatory portions. A very simple example occurs with a block-treatment classification if an additional “continuous” variable has been obtained: so, for instance, we might have

$$y_{ij} \doteq \mu + \beta_i + \tau_j + \gamma x_{ij},$$

where observations are indexed by ij , $\{\beta_i\}$ are block effects, $\{\tau_j\}$ are treatment effects, and x_{ij} is an observation on the ij th individual, such as initial weight in a growth feeding experiment. Continuous variables that are adjoined to a classificatory linear model as potential explanatory variables are given, for some quite obscure reason, the name *concomitant variables*, or *covariates* (see Section 4.13 and Chapter 8).

4.4 THE FITTING OF $\mathbf{y} \doteq \mathbf{X}\beta$

In formal terms the problem of fitting a model of the form (4.1) may be represented as follows: determine β and $\mathbf{X}\beta$ such that the badness of fit of \mathbf{y} by $\mathbf{X}\beta$, denoted by $BF(\mathbf{y}, \mathbf{X}\beta)$, is minimized. We shall not give a general discussion of this general problem. Instead we shall take, for our purposes,

$$BF(\mathbf{y}, \mathbf{X}\beta) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta), \quad (4.3)$$

and we refer to this as *least squares fitting*. We shall suppose at this point that β may be any vector in \mathbf{R}^p , that is, the components of $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$ may be any real numbers.

4.4.1 The Notion of Identifiability

Before proceeding with this, we ask a simple question: Suppose one is actually given the vector $\mathbf{X}\beta$ with knowledge of \mathbf{X} but not of β . Can one determine a linear function

$$\lambda'\beta = \lambda_1\beta_1 + \lambda_2\beta_2 + \dots + \lambda_p\beta_p$$

for given values $\lambda_1, \lambda_2, \dots, \lambda_p$? As background for the question, suppose we are given that, with $\beta_1 = \mu$, $\beta_2 = \alpha_1$, $\beta_3 = \alpha_2$,

$$\mu + \alpha_1 = 5$$

$$\mu + \alpha_2 = 7$$

can we determine α_1 ? This question can be thought about in very simple terms or in nonelementary but nonadvanced terms. Here is a geometric way of doing so.

We suppose $\mathbf{X}\beta = \mathbf{X}\beta_0$, that is, there is a vector β_0 which gives the vector $\mathbf{X}\beta$. Is then $\lambda'\beta$ necessarily determined to be $\lambda'\beta_0$? A slightly sophisticated way of thinking about this is to note the following: The equation $\mathbf{X}(\beta - \beta_0) = \mathbf{0}$ is equivalent to the vector $\beta - \beta_0$ being perpendicular to every vector that is the transpose of a row of the matrix \mathbf{X} . With

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \begin{pmatrix} \eta'_1 \\ \eta'_2 \\ \vdots \\ \eta'_n \end{pmatrix}$$

we say: The *row space* of \mathbf{X} is the set of all vectors $\{\sum_{i=1}^n a_i \eta'_i : a_i \text{'s arbitrary real numbers, and we denote this by } R = R(\mathbf{X})\}$. Also $\lambda'(\beta - \beta_0) = 0$ says that $\beta - \beta_0$ is

perpendicular to the vector

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_p \end{pmatrix},$$

which we write as $(\beta - \beta_0) \perp \lambda$. Then we can see that

$$(\beta - \beta_0) \perp R(X)$$

must imply

$$(\beta - \beta_0) \perp \lambda.$$

This happens if and only if $\lambda \in R(X)$ or $\lambda' = a'X$ for some vector a , or $\lambda = X'a$.

In the case of an approximative linear model, $y \doteq X\beta$, we say $\lambda'\beta$ is *identifiable* if $\lambda' = a'X$ for some vector a .

4.4.2 The Notion of Estimability

In the case of a stochastic linear model

$$y = X\beta + e$$

with $E(e) = 0$, where $E(\cdot)$ denotes the expectation or expected value, we say that $\lambda'\beta$ is linearly *estimable* if there exists a vector a such that

$$E(a'y) = \lambda'\beta$$

or because

$$\begin{aligned} E(a'y) &= E[a'(X\beta + e)] \\ &= E[a'X\beta + a'e] \\ &= a'X\beta + E(a'e) \\ &= a'X\beta \end{aligned}$$

we say, given that β is a completely free vector, that $\lambda'\beta$ is estimable if there exists a vector a such that

$$a'X = \lambda'$$

or, again, $\lambda \in R(X)$.

4.4.3 The Method of Least Squares

Now we proceed to the least squares fitting and give a sequence of results.

1. By differentiation of (4.3) with respect to the unknowns, $\beta_1, \beta_2, \dots, \beta_p$, we get the *normal equations* (NE)

$$X'Xb = X'y, \tag{4.4}$$

where we use b to denote the variable in the equations.

2. These equations are consistent for all $\mathbf{y} \in R^n$; that is, all vectors $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ where each y_i can be any real number. Hence solution vectors \mathbf{b} exist for any such vector \mathbf{y} .
3. Whatever solution we take, the vector \mathbf{Xb} is unique given a particular vector \mathbf{y} . This is so because for two solutions to (4.4), say \mathbf{b}_1 and \mathbf{b}_2 , we have $\mathbf{X}'\mathbf{Xb}_1 = \mathbf{X}'\mathbf{Xb}_2$ which implies $\mathbf{X}'\mathbf{X}(\mathbf{b}_1 - \mathbf{b}_2) = \mathbf{0}$ and hence

$$(\mathbf{b}_1 - \mathbf{b}_2)'\mathbf{X}'\mathbf{X}(\mathbf{b}_1 - \mathbf{b}_2) = 0$$

or

$$(\mathbf{Xb}_1 - \mathbf{Xb}_2)'(\mathbf{Xb}_1 - \mathbf{Xb}_2) = 0,$$

from which it follows that $\mathbf{Xb}_1 - \mathbf{Xb}_2 = \mathbf{0}$.

4. No matter what solution vector $\tilde{\mathbf{b}}$ we take, $BF(\mathbf{y}, \mathbf{X}\tilde{\mathbf{b}})$ is the minimum value of $(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$.
5. Because \mathbf{Xb} is unique, the NE necessarily give a unique answer for a given \mathbf{y} for any identifiable or estimable function $\lambda'\beta$. In fact, λ then is such that there exists a solution to the *conjugate normal equation*: $\mathbf{X}'\mathbf{X}\rho = \lambda$, and the fit for $\lambda'\beta$ is $\rho'\mathbf{X}'\mathbf{y}$.
6. There is, from (2) above, a vector \mathbf{b}_i such that

$$\mathbf{X}'\mathbf{Xb}_i = \mathbf{X}'\mathbf{e}_i,$$

where

$$\mathbf{e}'_i = (0 \cdots 0 \quad 1 \quad 0 \cdots 0) \quad (i = 1, 2, \dots, n).$$

\uparrow
*i*th position

So there is a matrix $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ such that

$$\mathbf{X}'\mathbf{XB} = \mathbf{X}'(\mathbf{e}_1 \quad \mathbf{e}_2 \cdots \mathbf{e}_n) = \mathbf{X}'\mathbf{I}_n = \mathbf{X}', \quad (4.5)$$

and \mathbf{XB} is unique.

7. From (4.5) we have

$$\mathbf{B}'\mathbf{X}'\mathbf{XB} = \mathbf{B}'\mathbf{X}'$$

and transposing yields

$$\mathbf{B}'\mathbf{X}'\mathbf{XB} = \mathbf{XB}.$$

So \mathbf{BX} is symmetric and idempotent (s.i.p.). This matrix \mathbf{XB} is determined solely by the matrix \mathbf{X} and we write

$$\mathbf{P}_\mathbf{X} = \mathbf{XB} \quad (4.6)$$

with

$$\mathbf{P}'_\mathbf{X} = \mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{X}^2, \quad (4.7)$$

these encompassing the symmetric idempotent properties. Furthermore

$$\mathbf{P}_\mathbf{X}\mathbf{X} = \mathbf{XBX} = \mathbf{B}'\mathbf{X}'\mathbf{X} = \mathbf{X}. \quad (4.8)$$

8. Premultiplying

$$X'Xb = X'y$$

by B' we get

$$B'X'Xb = B'X'y$$

or, using (4.6), (4.7), and (4.8),

$$Xb = P_X y.$$

9. We write

$$\begin{aligned} y &= P_X y + (I - P_X)y \\ &= a_X + a_0, \end{aligned} \quad (4.9)$$

where a_X represents the fit of y by Xb and a_0 represents the residual. Then

$$a_0' a_X = y'(I - P_X)' P_X y = 0.$$

So $P_X y$ and $(I - P_X)y$ are $n \times 1$ vectors that are perpendicular and

$$\begin{aligned} y'y &= (P_X y)'(P_X y) + [(I - P_X)y]'[(I - P_X)y] \\ &= y'P_X y + y'(I - P_X)y. \end{aligned} \quad (4.10)$$

10. This gives the very simple analysis of variance:

Explanatory Source	Degrees of Freedom	Sum of Squares
X	$\text{rank}(X)$	$y'P_X y$
Residual	$n - \text{rank}(X)$	$y'(I - P_X)y$
Total	n	$y'y$

11. We attach a number, the *degrees of freedom* (d.f.) associated with the explanatory source, X , equal to $\text{rank}(X)$, the rank of X , which is equal to the row rank of X or the column rank of X or the determinant rank of X . Also

$$\text{rank}(X) = \text{rank}(P_X)$$

because

$$P_X = XB, \quad \text{so } \text{rank}(P_X) \leq \text{rank}(X)$$

and

$$X = P_X X, \quad \text{so } \text{rank}(X) \leq \text{rank}(P_X).$$

A rationale for the number of d.f. is as follows: The fit for y is Xb for some b . Writing $Xb = b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$, we say that Xb is in the column space of X , denoted by $C(X)$. Now $C(X)$ is a space of dimension $r = \text{rank}(X)$. Similarly, the residual is $(I - P_X)y$ which is restricted to $C(I - P_X)$ which is a space of dimension $\text{rank}(I - P_X)$ which is equal to $n - r$, because $X'[(I - P_X)y] = 0$ for every y . So the elements of $(I - P_X)y$ are n linear forms which are restricted to be null in r ways.

12. We must emphasize that the matrix $\mathbf{P}_\mathbf{X}$ which is utterly intrinsic in the mathematics of the least squares approximation is not at all essential to the actual numerics of least squares fitting. We are given the NE: $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ and we have to find a solution, it does not matter which because $\mathbf{X}\mathbf{b}$ is invariant. We shall describe below how one may adjoin conditions on solutions in the form of $\mathbf{C}\mathbf{b} = \mathbf{c}$ (which we may take to be 0), so as to get a definite solution to the NE.
13. If $\tilde{\mathbf{b}}$ is a solution vector, we have

$$\mathbf{y}'\mathbf{P}_\mathbf{X}\mathbf{y} = (\mathbf{y}')(\mathbf{P}_\mathbf{X}\mathbf{y}) = \mathbf{y}'\mathbf{X}\tilde{\mathbf{b}} = \tilde{\mathbf{b}}'\mathbf{X}'\mathbf{y}. \quad (4.11)$$

The well-established phrase associated with this is:

The sum of squares removed by the fitting of \mathbf{y} by $\mathbf{X}\beta$ is equal to the sum of products of a solution of the NE and the right-hand sides of the NE, that is, the inner product of a solution vector and the right-hand side of the NE.

14. It is worth noting that $\text{rank}(\mathbf{X})$ is equal to the dimensionality of $C(\mathbf{X})$, that is with $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ so that \mathbf{x}_i is the i th column of \mathbf{X} , the set of all vectors

$$\left\{ \sum_{i=1}^p a_i \mathbf{x}_i : a_i \text{ a real number} \right\} \equiv C(\mathbf{X}).$$

Given the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$, we can find a maximal set of linearly independent vectors

$$\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_r\}$$

such that

$$\sum_{i=1}^r a_i \boldsymbol{\xi}_i = \mathbf{0} \quad \text{implies} \quad a_i = 0 \quad (i = 1, 2, \dots, r)$$

and every vector \mathbf{x}_i is given by

$$\mathbf{x}_i = \sum_{j=1}^r b_{ij} \boldsymbol{\xi}_j.$$

The number of vectors in such a maximal set is the column rank of \mathbf{X} .

We have seen that $\mathbf{P}_\mathbf{X} = \mathbf{X}\mathbf{B}$, $\mathbf{P}_\mathbf{X}\mathbf{y} = \mathbf{X}(\mathbf{B}\mathbf{y})$; so $\mathbf{P}_\mathbf{X}$ takes a vector \mathbf{y} into a vector that is in $C(\mathbf{X})$. Further,

$$(\mathbf{y} - \mathbf{P}_\mathbf{X}\mathbf{y})'\mathbf{P}_\mathbf{X}\mathbf{z} = \mathbf{y}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{P}_\mathbf{X}\mathbf{z} = 0$$

so $\mathbf{P}_\mathbf{X}$ projects \mathbf{y} orthogonally onto $C(\mathbf{X})$.

15. It is also clear that there exist maximal sets of independent row vectors in \mathbf{X} . One merely “works down” the row vectors keeping in a list those that are linearly independent of previous ones on that list. Necessarily, any such maximal set has

r members, where $r = \text{rank}(\mathbf{X})$. This tells us that there exists a set of linear functions of β , $\lambda'_1\beta$, $\lambda'_2\beta$, \dots , $\lambda'_r\beta$, such that with

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_r \end{pmatrix} = \begin{pmatrix} \lambda'_1 \\ \lambda'_2 \\ \vdots \\ \lambda'_r \end{pmatrix} \beta = \Lambda' \beta$$

it is the case that

$$\mathbf{X}\beta = \mathbf{Z}\Lambda'\beta = \mathbf{Z}\theta,$$

where \mathbf{Z} is $n \times r$ of rank r . Also, if $\lambda'\beta$ is identifiable or estimable then

$$\lambda'\beta = \nu'\theta$$

for some ν . In other words: a linear model $\mathbf{y} \doteq \mathbf{X}\beta$ of rank r on a parameter that is a p -vector ($p \geq r$, of course) can be written as a model $\mathbf{y} \doteq \mathbf{Z}\theta$ where \mathbf{Z} is $n \times r$ of rank r , and θ is a set of r linearly independent identifiable or estimable functions. This process is called *reparametrization* to full rank. Clearly, it can be done in many ways. We shall see that some ways are more natural or more convenient than others.

4.4.4 Theory of Linear Equations

We have seen that to obtain the fit for any identifiable or estimable function we merely have to obtain *any* solution of the NE: $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ and then if \mathbf{b}^* is any such solution, the solution for $\mathbf{X}\mathbf{b}$ is $\mathbf{X}\mathbf{b}^*$ and the solution for any identifiable function $\lambda'\beta$ is $\lambda'\mathbf{b}^*$. The question is, therefore, to exhibit ways of getting *one* solution.

1. We now give a few basic ideas on the theory of equations. A necessary and sufficient condition for the equations $\mathbf{A}\mathbf{x} = \mathbf{d}$ with unknown vector \mathbf{x} to be consistent can be expressed equivalently as
 - (a) $\mathbf{d} \in C(\mathbf{A})$ or
 - (b) $\text{rank}(\mathbf{A}|\mathbf{d}) = \text{rank}(\mathbf{A})$ or
 - (c) $\nu'\mathbf{A} = \mathbf{0}'$ implies $\nu'\mathbf{d} = 0$.
2. Notions of generalized inverses of matrices are useful. Consider the equation with a given real matrix \mathbf{A} in an unknown matrix \mathbf{X} ,

$$\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}. \quad (4.12)$$

It is solvable because of the following:

- (i) From basic matrix theory, there exist invertible \mathbf{P} and \mathbf{Q} such that

$$\mathbf{P}\mathbf{A}\mathbf{Q} = \begin{pmatrix} \mathbf{I}_r & \emptyset \\ \emptyset & \emptyset \end{pmatrix}$$

or

$$\mathbf{A} = \mathbf{P}^{-1} \begin{pmatrix} \mathbf{I}_r & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \mathbf{Q}^{-1}, \quad (4.13)$$

where $r = \text{rank}(\mathbf{A})$; and \emptyset are null-matrices of appropriate dimensions.

(ii) Take such \mathbf{P} and \mathbf{Q} . Then, take

$$\tilde{\mathbf{X}} = \mathbf{Q} \begin{pmatrix} \mathbf{I}_r & \mathbf{J} \\ \mathbf{K} & \mathbf{L} \end{pmatrix} \mathbf{P}, \quad (4.14)$$

where \mathbf{J} , \mathbf{K} , \mathbf{L} are arbitrary but of appropriate dimensions. It is easy to verify that $\mathbf{A}\tilde{\mathbf{X}}\mathbf{A} = \mathbf{A}$ and hence $\tilde{\mathbf{X}}$ of (4.14) is a solution and any solution is necessarily representable in this way.

(iii) Suppose \mathbf{A} is of rank r and the submatrix consisting of rows $\alpha_1, \alpha_2, \dots, \alpha_r$ and columns $\beta_1, \beta_2, \dots, \beta_r$ is $\tilde{\mathbf{A}}$ and is invertible. Then we can obtain an \mathbf{X} satisfying $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}$ by making up \mathbf{X} from $\tilde{\mathbf{A}}^{-1}$ by inserting this in rows $\beta_1, \beta_2, \dots, \beta_r$ and columns $\alpha_1, \alpha_2, \dots, \alpha_r$ and inserting zeros everywhere else (see Example 4.1).

Any solution \mathbf{X} of $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}$ is called a *generalized inverse* (or *g-inverse* for short) of \mathbf{A} and is usually denoted by \mathbf{A}^- , even though it is not unique.

3. Let \mathbf{A}^- be a particular generalized inverse of \mathbf{A} . Then

- (i) the equation $\mathbf{A}\mathbf{x} = \mathbf{d}$ is consistent if $\mathbf{A}\mathbf{A}^-\mathbf{d} = \mathbf{d}$,
- (ii) any solution to a consistent equation is of the form

$$\mathbf{x} = \mathbf{A}^-\mathbf{d} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z} \quad (4.15)$$

for some \mathbf{z} .

4. Hence, if we have to solve the NE, $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$, we can find an $(\mathbf{X}'\mathbf{X})^-$ by the procedure in (2) above and then take as solution

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}. \quad (4.16)$$

It is the case necessarily that

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{P}_\mathbf{X}$$

because with $\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{X}'$ and $\mathbf{P}_\mathbf{X} = \mathbf{X}\mathbf{B} = \mathbf{P}'_\mathbf{X} = \mathbf{P}_\mathbf{X}^2$ we have

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' = \mathbf{B}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{P}'_\mathbf{X}\mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{X}.$$

5. Another process for obtaining a solution to the consistent equation $\mathbf{A}\mathbf{x} = \mathbf{d}$ is merely to adjoin consistent equations $\mathbf{C}\mathbf{x} = \mathbf{c}$ so that the augmented equations have a unique solution. In our case, then, given the NE, $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ we adjoin conditions on solutions: $\mathbf{C}\mathbf{b} = \mathbf{c}$. It is the case that

- (i) If the equations in \mathbf{b} are to be consistent, then, from 1(c) above, we must have that $\nu'_1 \mathbf{X}'\mathbf{X} + \nu'_2 \mathbf{C} = \mathbf{0}'$ implies $\nu'_1 \mathbf{X}'\mathbf{y} + \nu'_2 \mathbf{c} = 0$. If we must have consistency for any conforming \mathbf{y} , we must have that $\nu'_1 \mathbf{X}'\mathbf{X} + \nu'_2 \mathbf{C} = \mathbf{0}'$ implies $\nu'_2 \mathbf{c} = 0$ and $\nu'_1 \mathbf{X}' = \mathbf{0}'$ which implies $\nu'_1 \mathbf{X}'\mathbf{X} = \mathbf{0}'$, $\nu'_2 \mathbf{c} = 0$. So, of course, $\mathbf{C}\mathbf{b} = \mathbf{c}$ must be consistent in \mathbf{b} .
- (ii) This condition on \mathbf{C} is, in slightly sophisticated terms, the condition

$$R(\mathbf{C}) \cap R(\mathbf{X}'\mathbf{X}) = \{\emptyset\}$$

or

$$R(\mathbf{X}) \cap R(\mathbf{C}) = \{\emptyset\}$$

because

$$\mathbf{B}'\mathbf{X}'\mathbf{X} = \mathbf{X}$$

implies

$$R(\mathbf{X}) \subset R(\mathbf{X}'\mathbf{X}),$$

which with

$$R(\mathbf{X}'\mathbf{X}) \subset R(\mathbf{X}),$$

gives

$$R(\mathbf{X}'\mathbf{X}) = R(\mathbf{X}).$$

Hence the prescription is clear: We adjoin to the NE the equations $\mathbf{C}\mathbf{b} = \mathbf{c}$, which are consistent (which we can accomplish merely by taking $\mathbf{c} = \mathbf{0}$) and which are such that the only identifiable or estimable function $\nu'\mathbf{C}\beta$ is the null function and such that $\text{rank}(\mathbf{C}) = p - r$, the column rank deficiency of \mathbf{X} . We shall give examples of this process later.

The method of obtaining a g -inverse as described in 2.(iii) above is used often in statistical software to obtain a solution to the NE (4.4). We shall return to this point in Section 6.11, but give a simple example here to illustrate the procedure.

EXAMPLE 4.1: For the model (4.2) with $b = t = 2$ we obtain $\mathbf{X}'\mathbf{X}$ in (4.4) as

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 4 & 2 & 2 & 2 & 2 \\ 2 & 2 & 0 & 1 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 & 0 \\ 2 & 1 & 1 & 0 & 2 \end{pmatrix} = \mathbf{A} \text{ say}$$

and, obviously, $\text{rank}(\mathbf{A}) = 3$. Then

$$\tilde{\mathbf{A}} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$$

with $\text{rank}(\tilde{\mathbf{A}}) = 3$ and

$$\tilde{\mathbf{A}}^{-1} = \begin{pmatrix} .75 & -.5 & -.5 \\ -.5 & 1 & 0 \\ -.5 & 0 & 1 \end{pmatrix}.$$

We then obtain

$$\mathbf{A}^- = (\mathbf{X}'\mathbf{X})^- \begin{pmatrix} .75 & -.5 & 0 & -.5 & 0 \\ -.5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -.5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

□

4.5 MOORE-PENROSE GENERALIZED INVERSE

We know from (4.5), (4.6) and (4.7) that there exists a \mathbf{U} such that

$$\mathbf{A}'\mathbf{A}\mathbf{U} = \mathbf{A}', \mathbf{P}_{\mathbf{A}} = \mathbf{A}\mathbf{U} = \mathbf{P}'_{\mathbf{A}} = \mathbf{P}_{\mathbf{A}}^2.$$

Similarly, there exists a \mathbf{V} such that

$$\mathbf{A}\mathbf{A}'\mathbf{V} = \mathbf{A}, \mathbf{A}'\mathbf{V} = \mathbf{P}_{\mathbf{A}'} = \mathbf{P}'_{\mathbf{A}'} = \mathbf{P}_{\mathbf{A}'}^2.$$

Now consider the matrix $\mathbf{A}^+ = \mathbf{V}'\mathbf{A}\mathbf{U}$, which is unique because $\mathbf{A}\mathbf{U}$ and $\mathbf{A}'\mathbf{V}$ are both unique. Clearly, this matrix \mathbf{A}^+ is determined uniquely by \mathbf{A} . It is an interesting matrix because

- (i) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$.
- (ii) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$.
- (iii) $\mathbf{A}\mathbf{A}^+$ is symmetric and idempotent.
- (iv) $\mathbf{A}^+\mathbf{A}$ is also symmetric and idempotent.

The properties (i)–(iv) above can be verified easily by making use repeatedly of (4.8) and the properties of $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{P}_{\mathbf{A}'}$ above:

- (i) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}\mathbf{V}'\mathbf{A}\mathbf{U}\mathbf{A} = \mathbf{A}\mathbf{U}\mathbf{A} = \mathbf{A}$.
- (ii) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{V}'\mathbf{A}\mathbf{U}\mathbf{A}\mathbf{V}'\mathbf{A}\mathbf{U} = \mathbf{V}'\mathbf{A}\mathbf{V}'\mathbf{A}\mathbf{U} = \mathbf{V}'\mathbf{A}\mathbf{U} = \mathbf{A}^+$.
- (iii) $(\mathbf{A}\mathbf{A}^+)' = (\mathbf{A}\mathbf{V}'\mathbf{A}\mathbf{U})' = (\mathbf{A}\mathbf{U})' = \mathbf{A}\mathbf{U}$
- (iv) $(\mathbf{A}^+\mathbf{A})' = (\mathbf{V}'\mathbf{A}\mathbf{U}\mathbf{A})' = (\mathbf{V}'\mathbf{A})' = \mathbf{V}'\mathbf{A}$.

Furthermore \mathbf{A}^+ is the unique solution to the equations in \mathbf{X}

$$\begin{aligned}\mathbf{A}\mathbf{X}\mathbf{A} &= \mathbf{A} \\ \mathbf{X}\mathbf{A}\mathbf{X} &= \mathbf{X} \\ (\mathbf{X}\mathbf{A})' &= \mathbf{X}\mathbf{A} \\ (\mathbf{A}\mathbf{X})' &= \mathbf{A}\mathbf{X}.\end{aligned}$$

This matrix is called the *Moore-Penrose generalized inverse* of \mathbf{A} , or, now briefly, the M-P inverse of \mathbf{A} . The M-P inverse has the following basic properties:

(i) $(\mathbf{A}')^+ = (\mathbf{A}^+)'$.

(ii) The shortest solution, that is, \mathbf{x} such that $\mathbf{x}'\mathbf{x}$ is minimized, of the consistent equation

$$\mathbf{A}\mathbf{x} = \mathbf{d}$$

is

$$\tilde{\mathbf{x}} = \mathbf{A}^+\mathbf{d}.$$

(iii) $(\mathbf{A}'\mathbf{A})^+ = \mathbf{A}^+(\mathbf{A}^+)'$

This gives us directly,

(iv) The shortest solution of the NE: $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ is

$$\tilde{\mathbf{b}} = \mathbf{X}^+\mathbf{y}. \quad (4.17)$$

(v) The vector perpendicular to the hyperplane set $\{\beta: \mathbf{C}\beta = \mathbf{c}\}$ is $\mathbf{C}^+\mathbf{c}$.

We see, of course, that \mathbf{A}^+ is a particular generalized inverse of \mathbf{A} .

4.6 CONDITIONED LINEAR MODEL

4.6.1 Affine Linear Model

Consider the model $\mathbf{y} = \mathbf{X}\beta$, with β not free in R^p but restricted by consistent equality conditions, $\mathbf{C}\beta = \mathbf{c}$, but otherwise free. Clearly we can write for any particular choice of \mathbf{C}^- ,

$$\beta = \mathbf{C}^-\mathbf{c} + (\mathbf{I} - \mathbf{C}^-\mathbf{C})\gamma \text{ for some } \gamma$$

or

$$\beta = \mathbf{C}^+\mathbf{c} + (\mathbf{I} - \mathbf{C}^+\mathbf{C})\gamma \text{ for some } \gamma.$$

So this restricted model is transformable to

$$\mathbf{y} = \mathbf{X}\mathbf{C}^-\mathbf{c} + \mathbf{X}(\mathbf{I} - \mathbf{C}^-\mathbf{C})\gamma$$

or

$$\mathbf{y} = \mathbf{X}\mathbf{C}^+\mathbf{c} + \mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})\gamma. \quad (4.18)$$

Table 4.1 Moore-Penrose Inverse

proc iml;				
A = {4 2 2 2 2,				
2 2 0 1 1,				
2 0 2 1 1,				
2 1 1 2 0,				
2 1 1 0 2};				
Ainv=ginv(A); print Ainv;				
MOORE-PENROSE INVERSE				
AINV				
0.0625	0.03125	0.03125	0.03125	0.03125
0.03125	0.265625	-0.234375	0.015625	0.015625
0.03125	-0.234375	0.265625	0.015625	0.015625
0.03125	0.015625	0.015625	0.265625	-0.234375
0.03125	0.015625	0.015625	-0.234375	0.265625

Either of these is appropriately called an *affine linear model*.

The fitting of this is really quite routine because with $\mathbf{y}^* = \mathbf{y} - \mathbf{X}\mathbf{C}^+\mathbf{c}$ the NE for $\boldsymbol{\gamma}$ in (4.18) is

$$[\mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})]'[\mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})]\tilde{\boldsymbol{\gamma}} = [\mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})]'\mathbf{y}^*$$

with shortest solution [see (4.17)]

$$\tilde{\boldsymbol{\gamma}} = [\mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})]^+[\mathbf{y} - \mathbf{X}\mathbf{C}^+\mathbf{c}].$$

Hence, as one may verify, the shortest solution to the least squares fitting of the whole problem is

$$\tilde{\mathbf{b}} = \mathbf{C}^+\mathbf{c} + [\mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})]^+[\mathbf{y} - \mathbf{X}\mathbf{C}^+\mathbf{c}]. \quad (4.19)$$

This mode of procedure is of considerable interest with regard to the mathematical structure that is being investigated. It does require, however, determination of M-P inverses and these are not at all easy to find, in general. Computer programs to do this exist, of course, such as SAS/IML (SAS Institute, Inc., 2002–2003).

The following example serves as an illustration.

EXAMPLE 4.2: Using the matrix **A** from Example 4.1 with the GINV function in SAS/IML yields, together with the input statement, the Moore-Penrose inverse given in Table 4.1. □

4.6.2 Normal Equations for the Conditioned Model

A quite different method of tackling the problem is to use Lagrange multipliers and differentiation. This gives the following equations, which we define to be the NE associated with $\mathbf{y} \doteq \mathbf{X}\beta$, $\mathbf{C}\beta = \mathbf{c}$,

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{C}' \\ \mathbf{C} & \emptyset \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{m} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{c} \end{pmatrix}, \quad (4.20)$$

in which the vector \mathbf{m} is a vector of undetermined multipliers. We now state basic properties of this NE (4.20):

- (i) It is consistent for all \mathbf{y} and all \mathbf{c} such that $\mathbf{C}\beta = \mathbf{c}$ is consistent.
- (ii) The vector $\mathbf{X}\beta$ is invariant over all solutions.
- (iii) Any part solution \mathbf{b} gives a minimum of $(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$ subject to $\mathbf{C}\beta = \mathbf{c}$.
- (iv) The minimum sum of squares is

$$\mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{b}'\mathbf{C}'\mathbf{m}.$$

Interesting and relevant properties of matrices involved are

$$(i) \quad \text{rank} \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{C}' \\ \mathbf{C} & \emptyset \end{pmatrix} = \text{rank} \begin{pmatrix} \mathbf{X} \\ \mathbf{C} \end{pmatrix} + \text{rank}(\mathbf{C}).$$

(ii) Suppose a generalized inverse of the coefficient matrix is given by

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{C}' \\ \mathbf{C} & \emptyset \end{pmatrix}^+ = \begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{F} & \mathbf{G} \end{pmatrix},$$

then a solution of (4.20) is

$$\begin{aligned} \mathbf{b} &= \mathbf{D}\mathbf{X}'\mathbf{y} + \mathbf{E}\mathbf{c} \\ \mathbf{m} &= \mathbf{F}\mathbf{X}'\mathbf{y} + \mathbf{G}\mathbf{c} \end{aligned}$$

and

$$\mathbf{X}\mathbf{b} = \mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{y} + \mathbf{X}\mathbf{E}\mathbf{c}.$$

Furthermore $\mathbf{X}\mathbf{D}\mathbf{X}'$ is symmetric idempotent and is, in fact, equal to

$$\mathbf{X}[\mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})]^+$$

[see (4.19)].

4.6.3 Different Types of Conditions

An interesting point arises if $C\beta$ is not identifiable or estimable. In that case $\nu'_1 X + \nu'_2 C = 0'$ implies $\nu'_1 X = 0'$, $\nu'_2 C = 0'$. It follows then from (4.20), which can be written in part as $(Xb - y)'X + m'C = 0'$, that $m'C = 0'$ and the minimum sum of squares is the same as that which would be obtained with just $y \doteq X\beta$, that is, without the restriction, $C\beta = c$. Also, we may note that if $R(C) \cap R(X) = \{\emptyset\}$, that is, no nontrivial $\nu'C\beta$ is identifiable or estimable, and $\text{rank } C = p - r$, the rank deficiency of X , then

$$\begin{pmatrix} X'X & C' \\ C & \emptyset \end{pmatrix}$$

is invertible and implies

$$(X'X + C'C)b = X'y + C'c \quad (4.21)$$

with solution

$$b = (X'X + C'C)^{-1}(X'y + C'c). \quad (4.22)$$

A particularly important special case of the preceding occurs when $C = \Lambda'X$ so that for every ν , it is the case that $\nu'C\beta$ is identifiable or estimable. In that case, in (4.20),

$$X'Xb + C'm = X'y$$

gives

$$X'Xb + X'\Lambda m = X'y$$

or

$$Xb = P_X(y - \Lambda m). \quad (4.23)$$

Then

$$c = Cb = \Lambda'Xb = \Lambda'P_X(y - \Lambda m)$$

so that

$$\Lambda'P_X\Lambda m = \Lambda'P_X y - c = \Lambda'P_X(y - X\beta_0) \quad (4.24)$$

for β_0 such that $C\beta_0 = c$. Then (4.24) gives a unique solution for $P_X\Lambda m$, which can be written as

$$P_X\Lambda m = P_X\Lambda(\Lambda'P_X\Lambda)^{-1}\Lambda'P_X(y - X\beta_0) \equiv Q(y - X\beta_0),$$

which, substituted into (4.23), gives

$$Xb = P_X y - Q(y - X\beta_0).$$

The minimum sum of squares is

$$(y - Xb)'(y - Xb) = y'(I - P_X)y + (y - X\beta_0)'Q(y - X\beta_0). \quad (4.25)$$

This is the sum of squares of residuals plus

$$(y - X\beta_0)'P_X\Lambda(\Lambda'P_X\Lambda)^{-1}\Lambda'P_X(y - X\beta_0).$$

But

$$\Lambda' \mathbf{P}_X \mathbf{y} = \Lambda' \mathbf{X} \tilde{\mathbf{b}} = \mathbf{C} \tilde{\mathbf{b}}, \quad \Lambda' \mathbf{P}_X \mathbf{X} \beta_0 = \Lambda' \mathbf{X} \beta_0 = \mathbf{c},$$

where $\Lambda' \tilde{\mathbf{b}}$ is the fit for $\Lambda' \beta$ in the model $\mathbf{y} \doteq \mathbf{X} \beta$, that is, the one without the restriction, and the additional sum of squares of deviation is then

$$(\mathbf{C} \tilde{\mathbf{b}} - \mathbf{c})' (\Lambda' \mathbf{P}_X \Lambda)^- (\mathbf{C} \tilde{\mathbf{b}} - \mathbf{c}).$$

Here, we may, of course, also use $(\Lambda' \mathbf{P}_X \Lambda)^+$ and if, as would normally be the case, $\mathbf{C} = \Lambda' \mathbf{X}$ is of full row rank, then $\Lambda' \mathbf{P}_X \Lambda$ is of full rank. Note that we can also write $\Lambda' \mathbf{P}_X \Lambda = \Lambda' \mathbf{X} (\mathbf{X}' \mathbf{X})^- \mathbf{X}' \Lambda = \mathbf{C} (\mathbf{X}' \mathbf{X})^- \mathbf{C}'$.

It is clear from the above that if we have the model: $\mathbf{y} \doteq \mathbf{X} \beta$, $\mathbf{C}_1 \beta = \mathbf{c}_1$, $\mathbf{C}_2 \beta = \mathbf{c}_2$, with $R(\mathbf{C}_1) \subset R(\mathbf{X})$, and $R(\mathbf{C}_2) \cap R(\mathbf{X}) = \{\emptyset\}$, then the space of possible fit vectors $\{\mathbf{X} \mathbf{b} : \mathbf{b} \in R^p\}$ is not restricted by the conditions $\mathbf{C}_2 \mathbf{b} = \mathbf{c}_2$. In other words, given \mathbf{b}_1 such that $\mathbf{C}_1 \mathbf{b}_1 = \mathbf{c}_1$, $\mathbf{C}_2 \mathbf{b}_1 = \mathbf{c}_2$, we can find \mathbf{b}_2 such that $\mathbf{C}_1 \mathbf{b}_2 = \mathbf{c}_1$, but $\mathbf{C}_2 \mathbf{b}_2 = \tilde{\mathbf{c}}_2 \neq \mathbf{c}_2$. Obviously, then the restriction $\mathbf{C}_2 \mathbf{b} = \mathbf{c}_2$ has no impact on the goodness of the representation of \mathbf{y} by a vector $\mathbf{X} \mathbf{b}$. More formally, with the model specified in the preceding, we have, with $\mathbf{C}_1 = \Lambda'_1 \mathbf{X}$, that the LS fit for $\mathbf{X} \beta$ is, according to (4.23),

$$\mathbf{X} \mathbf{b} = \mathbf{P}_X [\mathbf{y} - \Lambda_1 \mathbf{m}]$$

with \mathbf{m} satisfying (4.24), and it will be the case that

$$\{\mathbf{X} \mathbf{b} : \mathbf{C}_1 \mathbf{b} = \mathbf{c}_1, \mathbf{C}_2 \mathbf{b} = \mathbf{c}_2\} = \{\mathbf{X} \mathbf{b} : \mathbf{C}_1 \mathbf{b} = \mathbf{c}_1, \mathbf{C}_2 \mathbf{b} = \tilde{\mathbf{c}}_2, \tilde{\mathbf{c}}_2 \neq \mathbf{c}_2\}.$$

4.6.4 General Case

We can now describe very succinctly the situation with general conditions on parameters, $\mathbf{C} \beta = \mathbf{c}$. Necessarily, there exists with \mathbf{C} of dimensions $q \times p$, a partitioned matrix

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{pmatrix},$$

of order $q \times q$ and invertible, such that

$$\begin{aligned} R(\mathbf{T}_1 \mathbf{C}) &\subset R(\mathbf{X}) \\ R(\mathbf{T}_2 \mathbf{C}) \cap R(\mathbf{X}) &= \{\emptyset\} \end{aligned}$$

and the conditions $\mathbf{C} \beta = \mathbf{c}$ have impact on the fitting (whatever “reasonable” criterion of badness of fit is used) only with respect to the “portion”: $\mathbf{T}_1 \mathbf{C} \beta = \mathbf{T}_1 \mathbf{c}$. Furthermore, the conditions $\mathbf{T}_2 \mathbf{C} \beta = \mathbf{T}_2 \mathbf{c}$ have no impact on the resulting fit of $\mathbf{X} \beta$ by $\mathbf{X} \mathbf{b}$ for some vector \mathbf{b} .

A small final point merits a little consideration. How do we find \mathbf{T}_1 and \mathbf{T}_2 ? A possible procedure, not necessarily optimal with regard to computing, is as follows: Obtain first a basis for $R(\mathbf{X})$. Let the matrix of this be $\tilde{\mathbf{X}}$. Then consider

$$\begin{pmatrix} \tilde{\mathbf{X}} \\ \mathbf{C} \end{pmatrix}.$$

Reduce this to row echelon form. Then we shall obtain a matrix \tilde{C} as the lower nonnull part of the row echelon form. Then $C\beta = c$, $\tilde{C}\beta = \tilde{c}$, and $\tilde{C}\beta$ is not identifiable or estimable and the conditions $\tilde{C}\beta = \tilde{c}$ are not restrictive at all. The conditions $C\beta = c$ are equivalent then to $\tilde{C}\beta = \tilde{c}$ and $T_1 C\beta = T_1 c$. We are not giving here any attention to the question of how best to perform the operations that are involved.

4.7 TWO-PART LINEAR MODEL

4.7.1 Ordered Linear Models

Suppose we wish to consider explaining or approximating a variable vector y by means of a linear model on explanatory variables x_1, x_2, \dots, x_p , and z_1, z_2, \dots, z_q . Then we will wish to ask:

Do the variables z_1, z_2, \dots, z_q help in explaining y after we have used x_1, x_2, \dots, x_p ?
or

Do the variables x_1, x_2, \dots, x_p help in explaining y after we have used z_1, z_2, \dots, z_q ?
We may address these questions with matrix language by considering two *ordered* two-part linear models:

$$y \doteq X_1\beta_1 + X_2\beta_2 \quad (4.26)$$

and

$$y \doteq X_2\beta_2 + X_1\beta_1. \quad (4.27)$$

Here the *order* of writing is strongly relevant. To address this, taking the model (4.26) we can contemplate fitting y by $X_1\beta_1$, for which we are then looking at the residual $(I - P_{X_1})y$ [see (4.9)]. Then, multiplying (4.26) by $(I - P_{X_1})$, our model would give

$$(I - P_{X_1})y \doteq (I - P_{X_1})X_2\beta_2. \quad (4.28)$$

If now we fit (4.28) by the method of least squares we get the NE

$$X_2'(I - P_{X_1})X_2b_2 = X_2'(I - P_{X_1})y. \quad (4.29)$$

Rather naturally, we call (4.29) the *reduced normal equation*, RNE, associated with $X_2\beta_2$ after $X_1\beta_1$. Since (4.29) is a normal equation [that is, of the type $Z'Z\theta = Z'y$], it follows that $(I - P_{X_1})X_2b_2$ is uniquely determined. We achieve then the representation, based on (4.28),

$$y \doteq P_{X_1}y + (I - P_{X_1})X_2b_2. \quad (4.30)$$

In the same way, we have the RNE for $X_1\beta_1$, after $X_2\beta_2$, using model (4.27), which is

$$X_1'(I - P_{X_2})X_1b_1 = X_1'(I - P_{X_2})y. \quad (4.31)$$

The overall minimum sum of squares is then obtained from (4.30) as

$$\begin{aligned} & [(I - P_{X_1})y - (I - P_{X_1})X_2b_2]'[(I - P_{X_1})y - (I - P_{X_1})X_2b_2] \\ &= y'(I - P_{X_1})y - 2y'(I - P_{X_1})(I - P_{X_1})X_2b_2 + b_2'X_2'(I - P_{X_1})X_2b_2 \\ &= y'(I - P_{X_1})y - b_2'X_2'(I - P_{X_1})y. \end{aligned} \quad (4.32)$$

We note that in (4.32) $\mathbf{y}'(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{y}$ is the minimum sum of squares for fitting the model $\mathbf{y} \doteq \mathbf{X}_1\beta_1$. Hence (4.32) says that the additional sum of squares removed is the inner product of a solution vector of the RNE (4.29) and the right-hand vector of the RNE.

4.7.2 Using Orthogonal Projections

We shall now develop the situation by the use of \mathbf{P} -matrices, projection matrices. We use s.i.p. as an abbreviation for symmetric idempotent matrix.

We have the following:

- (i) There exists \mathbf{B}_1 such that $\mathbf{X}_1'\mathbf{X}_1\mathbf{B}_1 = \mathbf{X}_1'$ and $\mathbf{P}_{\mathbf{X}_1} = \mathbf{X}_1\mathbf{B}_1$ is s.i.p. with $\mathbf{P}_{\mathbf{X}_1}\mathbf{X}_1 = \mathbf{X}_1$ and $\mathbf{P}_{\mathbf{X}_1}$ is the orthogonal projector onto $C(\mathbf{X}_1)$, the column space of \mathbf{X}_1 .
- (ii) There exists \mathbf{B}_2 such that $\mathbf{X}_2'\mathbf{X}_2\mathbf{B}_2 = \mathbf{X}_2'$ and $\mathbf{P}_{\mathbf{X}_2} = \mathbf{X}_2\mathbf{B}_2$ is s.i.p. with $\mathbf{P}_{\mathbf{X}_2}\mathbf{X}_2 = \mathbf{X}_2$ and $\mathbf{P}_{\mathbf{X}_2}$ is the orthogonal projector onto $C(\mathbf{X}_2)$.
- (iii) There exist $\tilde{\mathbf{B}}_1, \tilde{\mathbf{B}}_2$ such that

$$(\mathbf{X}_1:\mathbf{X}_2)'(\mathbf{X}_1:\mathbf{X}_2) \begin{pmatrix} \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_2 \end{pmatrix} = (\mathbf{X}_1:\mathbf{X}_2)'$$

or

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \quad (4.33)$$

and

$$\mathbf{P}_{\mathbf{X}_1\mathbf{X}_2} = (\mathbf{X}_1:\mathbf{X}_2) \begin{pmatrix} \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_2 \end{pmatrix} = \mathbf{X}_1\tilde{\mathbf{B}}_1 + \mathbf{X}_2\tilde{\mathbf{B}}_2$$

is s.i.p. with $\mathbf{P}_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{X}_1:\mathbf{X}_2) = (\mathbf{X}_1:\mathbf{X}_2)$ and $\mathbf{P}_{\mathbf{X}_1\mathbf{X}_2}$ is the orthogonal projector onto $C(\mathbf{X}_1:\mathbf{X}_2)$.

For brevity of writing, we *now* use

$$\mathbf{P}_{\mathbf{X}_1} = \mathbf{P}_1, \quad \mathbf{P}_{\mathbf{X}_2} = \mathbf{P}_2, \quad \mathbf{P}_{\mathbf{X}_1\mathbf{X}_2} = \mathbf{P}_{12}.$$

Then

$$\mathbf{I} = \mathbf{P}_1 + (\mathbf{P}_{12} - \mathbf{P}_1) + (\mathbf{I} - \mathbf{P}_{12})$$

and

$$\begin{aligned} \mathbf{y} &= \mathbf{P}_1\mathbf{y} + (\mathbf{P}_{12} - \mathbf{P}_1)\mathbf{y} + (\mathbf{I} - \mathbf{P}_{12})\mathbf{y} \\ &\equiv \mathbf{a}_1 + \mathbf{a}_{2.1} + \mathbf{a}_{0.12} \quad (\text{by definition}). \end{aligned} \quad (4.34)$$

**Table 4.2 ($\mathbf{X}_2|\mathbf{X}_1$)—ANOVA for the Ordered
Model $\mathbf{y} \doteq \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$**

Explanatory Source	d.f.	Sum of Squares
$\mathbf{X}_1\boldsymbol{\beta}_1$	r_1	$\mathbf{y}'\mathbf{P}_1\mathbf{y}$
$\mathbf{X}_2\boldsymbol{\beta}_2$ after $\mathbf{X}_1\boldsymbol{\beta}_1$	$r_{12} - r_1$	$\mathbf{y}'(\mathbf{P}_{12} - \mathbf{P}_1)\mathbf{y}$
Residual	$n - r_{12}$	$\mathbf{y}'(\mathbf{I} - \mathbf{P}_{12})\mathbf{y}$
Total	n	$\mathbf{y}'\mathbf{y}$

Then we have decomposed the vector \mathbf{y} additively into three orthogonal vectors \mathbf{a}_1 , $\mathbf{a}_{2.1}$, and $\mathbf{a}_{0.12}$. We believe the reader will recognize immediately the rationale for the naming of these that we use. A little thought gives

$$\begin{aligned}
 \mathbf{P}_{12}\mathbf{P}_1 &= \mathbf{P}_1, \quad \text{so} \quad \mathbf{P}_1\mathbf{P}_{12} = \mathbf{P}_1 \quad \text{by transposition,} \\
 \mathbf{P}_1(\mathbf{I} - \mathbf{P}_{12}) &= \mathbf{0}, \\
 (\mathbf{P}_{12} - \mathbf{P}_1)' &= (\mathbf{P}_{12} - \mathbf{P}_1), \\
 (\mathbf{P}_{12} - \mathbf{P}_1)^2 &= (\mathbf{P}_{12} - \mathbf{P}_1), \\
 (\mathbf{I} - \mathbf{P}_{12})^2 &= (\mathbf{I} - \mathbf{P}_{12}).
 \end{aligned}$$

So \mathbf{P}_1 , $\mathbf{P}_{12} - \mathbf{P}_1$ and $\mathbf{I} - \mathbf{P}_{12}$ are each s.i.p. and we then have from (4.34)

$$\begin{aligned}
 \mathbf{y}'\mathbf{y} &= \mathbf{a}'_1\mathbf{a}_1 + \mathbf{a}'_{2.1}\mathbf{a}_{2.1} + \mathbf{a}'_{0.12}\mathbf{a}_{0.12} \\
 &= \mathbf{y}'\mathbf{P}_1\mathbf{y} + \mathbf{y}'(\mathbf{P}_{12} - \mathbf{P}_1)\mathbf{y} + \mathbf{y}'(\mathbf{I} - \mathbf{P}_{12})\mathbf{y}.
 \end{aligned}$$

This is nothing but the ANOVA corresponding to the *ordered* linear model: $\mathbf{y} \doteq \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$ which is normally presented as in Table 4.2 where $r_1 = \text{rank } \mathbf{X}_1$, $r_{12} = \text{rank}(\mathbf{X}_1;\mathbf{X}_2)$.

We see that the degrees of freedom associated with $\mathbf{y}'(\mathbf{P}_{12} - \mathbf{P}_1)\mathbf{y}$ are the degrees of freedom associated with $(\mathbf{P}_{12} - \mathbf{P}_1)\mathbf{y}$, which is $\text{rank}(\mathbf{P}_{12} - \mathbf{P}_1) = \text{trace}(\mathbf{P}_{12} - \mathbf{P}_1)$ (because $\mathbf{P}_{12} - \mathbf{P}_1$ is s.i.p.) $= \text{trace } \mathbf{P}_{12} - \text{trace } \mathbf{P}_1 = \text{rank}(\mathbf{P}_{12}) - \text{rank}(\mathbf{P}_1) = \text{rank}(\mathbf{X}_1;\mathbf{X}_2) - \text{rank}(\mathbf{X}_1)$.

We can abbreviate our naming of sources, clearly, to the following:

$$\begin{aligned}
 &\mathbf{X}_1 \\
 &\mathbf{X}_2|\mathbf{X}_1 \\
 &\mathbf{I}|\mathbf{X}_1\mathbf{X}_2
 \end{aligned}$$

a usage we shall find very handy.

Clearly, rather than fitting $\mathbf{X}_1\boldsymbol{\beta}_1$ first and then $\mathbf{X}_2\boldsymbol{\beta}_2$ to the residuals, we can fit $\mathbf{X}_2\boldsymbol{\beta}_2$ first and then $\mathbf{X}_1\boldsymbol{\beta}_1$ to the residuals. This corresponds to the identity

$$\mathbf{I} = \mathbf{P}_2 + (\mathbf{P}_{12} - \mathbf{P}_2) + (\mathbf{I} - \mathbf{P}_{12})$$

**Table 4.3 ($\mathbf{X}_1|\mathbf{X}_2$)—ANOVA for the
Ordered Model $\mathbf{y} \doteq \mathbf{X}_2\beta_2 + \mathbf{X}_1\beta_1$**

Explanatory Source	d.f.	Sum of Squares
\mathbf{X}_2	r_2	$\mathbf{y}'\mathbf{P}_2\mathbf{y}$
$\mathbf{X}_1 \mathbf{X}_2$	$r_{12} - r_2$	$\mathbf{y}'(\mathbf{P}_{12} - \mathbf{P}_2)\mathbf{y}$
$\mathbf{I} \mathbf{X}_2\mathbf{X}_1$	$n - r_{12}$	$\mathbf{y}'(\mathbf{I} - \mathbf{P}_{12})\mathbf{y}$
Total	n	$\mathbf{y}'\mathbf{y}$

and

$$\mathbf{y} = \mathbf{a}_2 + \mathbf{a}_{1 \cdot 2} + \mathbf{a}_{0 \cdot 12}.$$

We then get from this decomposition the ANOVA of Table 4.3.

We note that the residual sum of squares of Table 4.2 and 4.3, $\mathbf{y}'(\mathbf{I} - \mathbf{P}_{12})\mathbf{y}$, is the same as (4.32).

How do we use these ANOVAs? We suggest that this is rather obvious. If we need $\mathbf{X}_2\beta_2$ after $\mathbf{X}_1\beta_1$, then it is the case that

$$\frac{SS(\mathbf{X}_2|\mathbf{X}_1)}{SS(\mathbf{I}|\mathbf{X}_1\mathbf{X}_2)}$$

is “large.” Just what we mean here by “large” will be clarified later (see Section 4.17).

4.7.3 Orthogonal ANOVA

A small question is “When are the two ANOVAs the same apart from naming?” This happens only if

$$\mathbf{y}'(\mathbf{P}_{12} - \mathbf{P}_1)\mathbf{y} = \mathbf{y}'\mathbf{P}_2\mathbf{y}, \quad \text{for all } \mathbf{y}$$

and

$$\mathbf{y}'(\mathbf{P}_{12} - \mathbf{P}_2)\mathbf{y} = \mathbf{y}'\mathbf{P}_1\mathbf{y}, \quad \text{for all } \mathbf{y}$$

or

$$\mathbf{P}_{12} = \mathbf{P}_1 + \mathbf{P}_2$$

or

$$\mathbf{X}_1'\mathbf{P}_{12}\mathbf{X}_2 = \mathbf{X}_1'\mathbf{P}_1\mathbf{X}_2 + \mathbf{X}_1'\mathbf{P}_2\mathbf{X}_2$$

or, using the properties of \mathbf{P}_{12} , \mathbf{P}_1 , \mathbf{P}_2 ,

$$\mathbf{X}_1'\mathbf{X}_2 = \mathbf{X}_1'\mathbf{X}_2 + \mathbf{X}_1'\mathbf{X}_2$$

or

$$\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}.$$

Conversely, if $\mathbf{X}_1' \mathbf{X}_2 = \emptyset$, it is clear from (4.33) that

$$\mathbf{P}_{12} = \mathbf{P}_1 + \mathbf{P}_2$$

and we have just one ANOVA (see also Section 4.11). This is referred to as *orthogonal* ANOVA.

Little addenda to the above are as follows:

- (a) $\mathbf{P}_{12} - \mathbf{P}_1$ is the orthogonal projector onto $C[(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_2]$, and
- (b) $\mathbf{P}_{12} - \mathbf{P}_2$ is the orthogonal projector onto $C[(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1]$

4.8 SPECIAL CASE OF A PARTITIONED MODEL

Consider the model: $\mathbf{y} \doteq \mathcal{J}\mu + \mathbf{X}\beta$, where $\mathcal{J}' = (1, 1, \dots, 1)$ with n components. Then we see that

$$\mathbf{P}_{\mathcal{J}} = \frac{1}{n} \mathcal{J}\mathcal{J}'. \quad (4.35)$$

The RNE for $\mathbf{X}\beta$ is

$$\mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathcal{J}\mathcal{J}' \right) \mathbf{X} \mathbf{b} = \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathcal{J}\mathcal{J}' \right) \mathbf{y}. \quad (4.36)$$

We obtain (4.36) merely by performing least squares fitting on \mathbf{y} with the mean \bar{y} subtracted from each y and with the mean of each column of \mathbf{X} subtracted from the elements of that column.

The sum of squares for the explanatory source $\mathcal{J}\mu$ is $\mathbf{y}' \mathbf{P}_{\mathcal{J}} \mathbf{y}$, which is nothing but $\frac{1}{n} \mathbf{y}' \mathcal{J}\mathcal{J}' \mathbf{y}$, or $\frac{1}{n}$ [square of total of y], a quantity commonly called the *correction factor* in ANOVAs “around the mean.”

4.9 THREE-PART MODELS

Here we consider

$$\mathbf{y} \doteq \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{X}_3 \beta_3$$

with six ordered three-part linear models, one for each of the six orderings of \mathbf{X}_1 , \mathbf{X}_2 , and \mathbf{X}_3 . We have thus six ANOVAs, which we represent in Table 4.4. To explain what the associated sums of squares are we use $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_{12}, \mathbf{P}_{13}, \mathbf{P}_{23}$ and \mathbf{P}_{123} , extending in an obvious way the results of Section 4.7, and merely give an example:

$$SS(\mathbf{X}_2 | \mathbf{X}_1 \mathbf{X}_3) = \mathbf{y}' (\mathbf{P}_{123} - \mathbf{P}_{13}) \mathbf{y}.$$

It is rather easy to see that the matrix of any sum of squares is s.i.p. The third ANOVA of Table 4.4 corresponds to the identity

$$\mathbf{I} = \mathbf{P}_2 + (\mathbf{P}_{12} - \mathbf{P}_2) + (\mathbf{P}_{123} - \mathbf{P}_{12}) + (\mathbf{I} - \mathbf{P}_{123}),$$

Table 4.4 The 6 ANOVAs with an Unordered Three-Part Model

Explanatory Sources					
\mathbf{X}_1	\mathbf{X}_1	\mathbf{X}_2	\mathbf{X}_2	\mathbf{X}_3	\mathbf{X}_3
$\mathbf{X}_2 \mathbf{X}_1$	$\mathbf{X}_3 \mathbf{X}_1$	$\mathbf{X}_1 \mathbf{X}_2$	$\mathbf{X}_3 \mathbf{X}_2$	$\mathbf{X}_1 \mathbf{X}_3$	$\mathbf{X}_2 \mathbf{X}_3$
$\mathbf{X}_3 \mathbf{X}_1\mathbf{X}_2$	$\mathbf{X}_2 \mathbf{X}_1\mathbf{X}_3$	$\mathbf{X}_3 \mathbf{X}_1\mathbf{X}_2$	$\mathbf{X}_1 \mathbf{X}_2\mathbf{X}_3$	$\mathbf{X}_2 \mathbf{X}_1\mathbf{X}_3$	$\mathbf{X}_1 \mathbf{X}_2\mathbf{X}_3$
$\mathbf{I} \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$	$\mathbf{I} \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$	$\mathbf{I} \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$	$\mathbf{I} \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$	$\mathbf{I} \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$	$\mathbf{I} \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$

giving

$$\mathbf{y} = \mathbf{a}_2 + \mathbf{a}_{1.2} + \mathbf{a}_{3.12} + \mathbf{a}_{0.123},$$

that is, the vector \mathbf{y} has been decomposed additively into four vectors that are orthogonal.

4.10 TWO-WAY CLASSIFICATION WITHOUT INTERACTION

We have given earlier [see (4.2)] the data model structure

$$\mathbf{y} = \mathcal{J}\mu + \mathbf{X}_r\mathbf{r} + \mathbf{X}_c\mathbf{c}, \quad (4.37)$$

where \mathbf{X}_r and \mathbf{X}_c are incidence matrices with $\mathbf{X}_r\mathcal{J} = \mathcal{J}$, $\mathbf{X}_c\mathcal{J} = \mathcal{J}$, the \mathcal{J} -vectors having appropriate dimensions. This represents a special case of a three-part model, but a very important one in the context of comparative experiments (see Chapters 9 and II.1). In this case we are interested, as we shall see, only in two ANOVAs. We are interested in the extent to which incorporation of $\mathbf{X}_r\mathbf{r}$ and/or $\mathbf{X}_c\mathbf{c}$ improve approximation of \mathbf{y} by $\mathcal{J}\mu$. We then have only two relevant ANOVAs:

$$\begin{array}{c} \mathcal{J} \\ \mathbf{X}_r|\mathcal{J} \\ \mathbf{X}_c|\mathcal{J}\mathbf{X}_r \\ \mathbf{I}|\mathcal{J}\mathbf{X}_r\mathbf{X}_c \end{array} \quad \text{and} \quad \begin{array}{c} \mathcal{J} \\ \mathbf{X}_c|\mathcal{J} \\ \mathbf{X}_r|\mathcal{J}\mathbf{X}_c \\ \mathbf{I}|\mathcal{J}\mathbf{X}_r\mathbf{X}_c \end{array}.$$

A very natural question is whether we shall have

$$SS(\mathbf{X}_c|\mathcal{J}\mathbf{X}_r) = SS(\mathbf{X}_c|\mathcal{J}),$$

which then induces and is equivalent to

$$SS(\mathbf{X}_r|\mathcal{J}\mathbf{X}_c) = SS(\mathbf{X}_r|\mathcal{J}).$$

It is clear from Section 4.7 that if this is to happen we must have

$$\mathbf{y}'(\mathbf{P}_{rc} - \mathbf{P}_r)\mathbf{y} = \mathbf{y}'(\mathbf{P}_c - \mathbf{P}_{\mathcal{J}})\mathbf{y},$$

where \mathbf{P}_J is the orthogonal projector onto $C(J)$, \mathbf{P}_r is the orthogonal projector onto $C(\mathbf{X}_r) = C(J:\mathbf{X}_r)$, \mathbf{P}_c is the orthogonal projector onto $C(\mathbf{X}_r:\mathbf{X}_c) = C(J:\mathbf{X}_r:\mathbf{X}_c)$. The equality must hold for all \mathbf{y} so we need

$$\mathbf{P}_{rc} - \mathbf{P}_r = \mathbf{P}_c - \mathbf{P}_J.$$

This gives

$$\mathbf{X}'_r(\mathbf{P}_{rc} - \mathbf{P}_r)\mathbf{X}_c = \mathbf{X}'_r(\mathbf{P}_c - \mathbf{P}_J)\mathbf{X}_c$$

or

$$\emptyset = \mathbf{X}'_r\mathbf{X}_c - \mathbf{X}'_r\mathbf{P}_J\mathbf{X}_c$$

or, using (4.35),

$$\mathbf{X}'_r\mathbf{X}_c = \frac{1}{n}(\mathbf{X}'_rJ)(J'\mathbf{X}_c). \quad (4.38)$$

Since the elements of $\mathbf{X}'_r\mathbf{X}_c$ are the numbers of observations for each row-column combination, (4.38) tells us that we have one ANOVA if the frequencies are proportional. Contrariwise, if $\mathbf{X}'_r\mathbf{X}_c = \mathbf{X}'_r\mathbf{P}_J\mathbf{X}_c$, then with $\mathbf{P}_{rc} = \mathbf{X}_r\tilde{\mathbf{B}}_r + \mathbf{X}_c\tilde{\mathbf{B}}_c$, where

$$\begin{pmatrix} \mathbf{X}'_r\mathbf{X}_r & \mathbf{X}'_r\mathbf{X}_c \\ \mathbf{X}'_c\mathbf{X}_r & \mathbf{X}'_c\mathbf{X}_c \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{B}}_r \\ \tilde{\mathbf{B}}_c \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_r \\ \mathbf{X}'_c \end{pmatrix} \quad (4.39)$$

with $\mathbf{X}'_r\mathbf{X}_r\mathbf{B}_r = \mathbf{X}'_r$, $\mathbf{X}'_c\mathbf{X}_c\mathbf{B}_c = \mathbf{X}'_c$, $\mathbf{P}_r = \mathbf{X}_r\mathbf{B}_r$, $\mathbf{P}_c = \mathbf{X}_c\mathbf{B}_c$, we obtain from (4.39), premultiplying by \mathbf{B}'_r and \mathbf{B}'_c , respectively,

$$\mathbf{X}_r\tilde{\mathbf{B}}_r + \mathbf{P}_r\mathbf{P}_J\mathbf{X}_c\tilde{\mathbf{B}}_c = \mathbf{P}_r \quad (4.40)$$

$$\mathbf{P}_c\mathbf{P}_J\mathbf{X}_r\tilde{\mathbf{B}}_r + \mathbf{X}_c\tilde{\mathbf{B}}_c = \mathbf{P}_c. \quad (4.41)$$

But

$$\mathbf{P}_r\mathbf{P}_J = \mathbf{P}_J, \mathbf{P}_c\mathbf{P}_J = \mathbf{P}_J$$

so, adding (4.40) and (4.41),

$$\mathbf{P}_{rc} + \mathbf{P}_J\mathbf{P}_{rc} = \mathbf{P}_r + \mathbf{P}_c.$$

Since $\mathbf{P}_{rc}\mathbf{P}_J = \mathbf{P}_J$, this implies

$$\mathbf{P}_{rc} + \mathbf{P}_J = \mathbf{P}_r + \mathbf{P}_c$$

or

$$\mathbf{P}_{rc} - \mathbf{P}_c = \mathbf{P}_r - \mathbf{P}_J$$

so that we have just one ANOVA. Hence, proportional frequencies are a necessary and sufficient condition for both ANOVAs to be identical.

4.11 K-PART LINEAR MODEL

4.11.1 The General Model and Its Sums of Squares

Consider the linear model

$$\mathbf{y} \doteq \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \cdots + \mathbf{X}_k\beta_k.$$

Clearly, there are $k!$ associated ordered k -part linear models. Taking the original order, we have

$$\mathbf{P}_1, \mathbf{P}_{12}, \mathbf{P}_{123}, \dots, \mathbf{P}_{12\dots k}$$

and

$$\mathbf{I} = \mathbf{P}_1 + (\mathbf{P}_{12} - \mathbf{P}_1) + \cdots + (\mathbf{P}_{12\dots k} - \mathbf{P}_{12\dots \overline{k-1}}) + (\mathbf{I} - \mathbf{P}_{12\dots k}),$$

giving

$$\mathbf{y} = \mathbf{a}_1 + \mathbf{a}_{2\cdot 1} + \cdots + \mathbf{a}_{k\cdot 12\dots \overline{k-1}} + \mathbf{a}_{0\cdot 12\dots k}$$

and

$$\mathbf{y}'\mathbf{y} = \mathbf{a}'_1\mathbf{a}_1 + \mathbf{a}'_{2\cdot 1}\mathbf{a}_{2\cdot 1} + \cdots + \mathbf{a}'_{k\cdot 12\dots \overline{k-1}}\mathbf{a}_{k\cdot 12\dots \overline{k-1}} + \mathbf{a}'_{0\cdot 12\dots k}\mathbf{a}_{0\cdot 12\dots k}.$$

This may be presented as an ANOVA:

Explanatory Source
\mathbf{X}_1
$\mathbf{X}_2 \mathbf{X}_1$
$\mathbf{X}_3 \mathbf{X}_1\mathbf{X}_2$
\vdots
$\mathbf{X}_k \mathbf{X}_1\mathbf{X}_2\dots\mathbf{X}_{k-1}$
$\mathbf{I} \mathbf{X}_1\mathbf{X}_2\dots\mathbf{X}_k$
\mathbf{I}

in which degrees of freedom and sums of squares may be written down at sight. Every sum of squares is obtainable by considering for some $\mathbf{Z}_1, \mathbf{Z}_2$

$$\begin{aligned}\mathbf{y} &\doteq \mathbf{Z}_1\gamma_1 \\ \mathbf{y} &\doteq \mathbf{Z}_1\gamma_1 + \mathbf{Z}_2\gamma_2\end{aligned}$$

with respective sums of squares

$$\gamma_1^*'\mathbf{Z}_1\mathbf{y}, \quad (4.42)$$

where $\mathbf{Z}_1'\mathbf{Z}_1\gamma_1^* = \mathbf{Z}_1'\mathbf{y}$, and

$$\tilde{\gamma}_1'\mathbf{Z}_1'\mathbf{y} + \tilde{\gamma}_2'\mathbf{Z}_2'\mathbf{y}, \quad (4.43)$$

where

$$(\mathbf{Z}_1:\mathbf{Z}_2)'(\mathbf{Z}_1:\mathbf{Z}_2) \begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} = (\mathbf{Z}_1:\mathbf{Z}_2)'\mathbf{y}$$

and then taking (4.43) minus (4.42). For example, to obtain $SS(\mathbf{X}_3|\mathbf{X}_1\mathbf{X}_2)$ we take

$$\mathbf{Z}_1 = (\mathbf{X}_1:\mathbf{X}_2), \quad \gamma_1 = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

and

$$\mathbf{Z}_2 = \mathbf{X}_3, \quad \gamma_2 = \beta_3.$$

The difference (4.43) minus (4.42) can then be expressed as

$$SS(\mathbf{Z}_2|\mathbf{Z}_1) = SS(\mathbf{Z}_1\mathbf{Z}_2) - SS(\mathbf{Z}_1)$$

or, alternatively, as

$$SS(\mathbf{X}_3|\mathbf{X}_1\mathbf{X}_2) = SS(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3) - SS(\mathbf{X}_1\mathbf{X}_2).$$

This shows that the sums of squares are obtained by fitting sequentially “larger” models, that is,

$$\mathbf{y} \doteq \mathbf{X}_1\beta_1$$

then

$$\mathbf{y} \doteq \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$$

and then

$$\mathbf{y} \doteq \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{X}_3\beta_3$$

etc., and obtain $SS(\mathbf{X}_1)$, $SS(\mathbf{X}_1\mathbf{X}_2)$, $SS(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3)$ etc., respectively, and then

$$SS(\mathbf{X}_2|\mathbf{X}_1) = SS(\mathbf{X}_1\mathbf{X}_2) - SS(\mathbf{X}_1)$$

and

$$SS(\mathbf{X}_3|\mathbf{X}_1\mathbf{X}_2) = SS(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3) - SS(\mathbf{X}_1\mathbf{X}_2)$$

etc. It is for this reason that the sums of squares in the table above are often referred to as *sequential sums of squares* in the context of the k -part linear model. We emphasize again, that the use of \mathbf{P} -matrices is particularly valuable for mathematical purposes.

Among the $k!$ orderings for the k -part linear model we can identify k orderings such that for the i th ordering $\mathbf{X}_i\beta_i$ occurs in the last position ($i = 1, 2, \dots, k$). Then the last sum of squares of the sequential sums of squares is

$$SS(\mathbf{X}_i|\text{all other } \mathbf{X}_j) = SS(\mathbf{X}_1\mathbf{X}_2 \dots \mathbf{X}_k) - SS(\text{all } \mathbf{X}_j \text{ except } \mathbf{X}_i)$$

for $i = 1, 2, \dots, k$. These k sums of squares are referred to as *partial sums of squares*. These play an important role for nonorthogonal models, that is, models for which conditions corresponding to those given in Section 4.7.3 do not hold, for testing hypotheses involving $\mathbf{X}_i\beta_i$ (see, for instance, Sections 8.3.5, 8.8, 9.8, 9.10, and 13.4).

For later reference we mention here that for the Statistical Analysis System (SAS) package (SAS Institute, Inc. 2002-2003) the sequential sums of squares correspond to the Type I sums of squares and the partial sums of squares correspond to the Type III sums of squares.

4.11.2 The Means Model

It is relevant to give a little detail of alternative modes of specification of models.

The simple case which we have denoted by $\mathbf{y} \doteq \mathcal{J}\mu + \mathbf{X}_1\beta$ (see Section 4.8) where \mathbf{X}_1 is a $n \times p$ binary (0, 1) matrix, can be written as $\mathbf{y} \doteq \mathbf{X}_1\gamma$, because $\mathcal{J} = \mathcal{J}_n = \mathbf{X}_1\mathcal{J}_p$, so $\gamma = \beta + \mathcal{J}_p\mu$. When written in the latter mode, the rank of the model equals rank (\mathbf{X}_1) , and there are no dependences that the model $\mathbf{y} = \mathcal{J}\mu + \mathbf{X}_1\beta$ contains. This model without dependences is called a *cell means model* (e.g., Hocking 1985, 2003) or *means model* for short, which we shall write as $\mathbf{y} = \mathbf{X}\mu$.

In the case of the two-way classification without interaction, we have written the model as $\mathbf{y} \doteq \mathcal{J}\mu + \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$. We use this because it leads, automatically and directly, to the two relevant and interesting ANOVAs. Alternatively, we may write this in the form of a means model as

$$y_{ijk} \doteq \mu_{ij} \quad (4.44)$$

with conditions on the model as

$$\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..} = 0 \quad (4.45)$$

for all i, j , or, as

$$\mu_{ij} - \mu_{ij'} - \mu_{i'j} + \mu_{i'j'} = 0 \quad (4.46)$$

for all i, i', j, j' .

We have given earlier, modes of presentation of a conditional linear model. To write either of the alternative models in matrix form and then to construct the normal equations is an unpleasant chore. Additionally, to go from the conditional means model, say

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mu \\ \mathbf{C}\mu &= \mathbf{0} \end{aligned}$$

with \mathbf{C} of the form appropriate to (4.45) or (4.46) to consideration of what linear forms in β_1 and in β_2 are identifiable is awkward.

We want, of course, to write models for higher classification data structures. Consider a three-way structure with observations indicated by i = level of factor 1, j = level of factor 2, k = level of factor 3, and l = level of observation within levels i, j, k of factors 1, 2, and 3. We prefer the model written as

$$\mathbf{y} \doteq \mathcal{J}\mu + \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{X}_3\beta_3 + \mathbf{X}_{12}\beta_{12} + \mathbf{X}_{13}\beta_{13} + \mathbf{X}_{23}\beta_{23} + \mathbf{X}_{123}\beta_{123},$$

in which we include the possibility of interactions between factors 1, 2, and 3. How are we to represent the case of no triple interaction? The only alternative way is to write the means model as

$$y_{ijkl} \doteq \mu_{ijk} \quad (4.47)$$

and then adjoin conditions, such as

$$\mu_{ijk} - \mu_{ijk'} - \mu_{ij'k} - \mu_{ij'k} + \mu_{i'j'k} + \mu_{ij'k'} + \mu_{i'jk'} - \mu_{i'j'k'} = 0 \quad (4.48)$$

for all i, i', j, j', k, k' . To write the model with no interactions we have to write (4.47) and (4.48) with, also,

$$\begin{aligned}\mu_{ijk} - \mu_{ijk'} - \mu_{ij'k} + \mu_{ij'k'} &= 0, & \text{for all } i, j, j', k, k' \\ \mu_{ijk} - \mu_{i'jk} - \mu_{ijk'} + \mu_{i'jk'} &= 0, & \text{for all } i, i', j, k, k'\end{aligned}$$

and

$$\mu_{ijk} - \mu_{i'jk} - \mu_{ij'k} + \mu_{i'j'k} = 0, \quad \text{for all } i, i', j, j', k.$$

To write this as

$$\mathbf{y} = \mathbf{X}^* \boldsymbol{\mu}$$

with conditions on $\boldsymbol{\mu}$, given by $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ is exceedingly awkward and space filling.

We suggest, by way of this brief discussion, that using means models is, in general, not good data analysis procedure. Such models do not, without additional notation and commentary, reflect the structure of the data and hence are not suggestive concerning tests of hypothesis or estimation of linear functions of interest to the researcher. For example, model (4.44) can just as well represent a data structure where factor 2 is nested within factor 1. Similarly, model (4.47) could represent, among several possibilities, a data structure where factor 3 is nested in factor 2, and factors 1 and 2 are crossed. Such ambiguity can lead easily to inappropriate inference and hence to erroneous conclusions and explanations of the data (see also Section 4.12.7), although the proponents of the means model argue contrariwise (Hocking and Speed, 1975).

Models of the form we prefer are said to be *overparameterized models*. We discuss such models in more detail in Section 4.12.7. We shall show that for balanced classificatory data structures the perceived disadvantage of overparameterized models can be dealt with very easily. It is only for certain types of unbalanced data structures that the means model seems preferable (see Section 4.13.2).

4.12 BALANCED CLASSIFICATORY STRUCTURES AND ANALYSIS OF VARIANCE

In the previous sections we have considered linear models from a rather general point of view without distinguishing between what are usually referred to as regression models and classificatory or classification models. The reason, of course, is that the mathematics associated with these different models is really the same. In the context of comparative experiments, however, classificatory models play a major role. It is useful then to digress briefly from the general discussion and introduce and illustrate some important concepts for classificatory models. More specifically we want to discuss certain data structures, how such structures lead to classificatory linear models, and associated analyses of variance.

4.12.1 Factors, Levels, and Partitions

The analysis of variance (ANOVA) was first developed for what we call *balanced classificatory structures*. We have a set of individuals which can be partitioned or classified by one or more factors of classification. So human individuals can be classified by country of birth, by sex, by religion of ancestors, and so on. A *factor of classification* thus gives a partition of the set of individuals into disjoint subclasses. We denote a factor of classification by a capital letter, say, A , and we call the subsets produced by the factor the *levels of the factor*. So with any one factor every individual in the set possesses a level of each factor. Obviously, the set can be partitioned into subclasses by whatever number of factors is given by the data. So with two factors, which we denote by A and B , every individual in the whole set possesses a level of each of the two factors. There is a critically important attribute of the relationship between two factors or two subsets of the totality of factors. Clearly, a combination of factors is itself a factor because, for example, we can partition the set of individuals by two factors, A and B , and every individual is at a particular level of factor A and a particular level of factor B , so every individual is at a particular level of the joint factor AB . So then with four factors A , B , C , and D , we have four single factors, A , B , C , and D and we have six factors involving two factors, AB , AC , AD , BC , BD , and CD , and we have four factors involving three factors, BCD , ACD , ABD , and ABC , with one factor involving all four simple factors which we denote, naturally, by $ABCD$. We call the constituent factors in any joint factor by the name of the single factors.

4.12.2 Nested, Crossed, and Confounded Factors

Now consider two factors. Let these be sex, S , and continent of birth, C . Then there are individuals in every subclass formed by the joint factor, SC . In contrast, let the two factors be county of residence and state of residence which we denote by C and S respectively. For exposition we assume that the states and counties of the USA have different names, which is not “quite true.” Then there is only one county called Story County and one state called Iowa, and Story County is in the State of Iowa. It is obvious then that all individuals in Story County, having the same level of the factor county, have the same level of state, namely Iowa. In this case we say that the factor C is *nested* by the factor S , or, equivalently, that the factor S *nects* the factor C . Clearly, in this situation, we cannot have county Story in the state of Virginia so the combination (Story, Virginia) is not a possible combination.

It is clear that in this case, the joint factor CS gives the same partition as the factor C . The same type of relationship can be considered with joint factors, say AB and CD . So we can say, for example, that the joint factor AB nests the factor C , or that the joint factor AB nests the joint factor CD . And so on.

If two factors A and B or two joint factors, say, AB and CD are such that one does not nest the other, they are said to be *crossed*.

There is a third type of relationship between two factors. Suppose we have t EUs and t treatments, and that we assign each of the treatments to one unit. Then it is clear that the factors, units and treatments, are not crossed and that one is not nested by the other. We say that units and treatments are *completely confounded*. We say this

because whatever difference we observe between the units receiving, say, treatment 1 and treatment 2 can be explained by the difference caused by the difference between the treatments, or by the difference between the units on which treatment 1 and treatment 2 fell.

4.12.3 The Notion of Balance

We now turn to the matter of balance. We have a data set consisting of a set of individuals which may be humans, mice, pieces of engineering equipment, for example, which can be classified by a factor A . If the number of individuals at each level of A is the same for all levels, we say that the data set is *balanced* with respect to factor A . If we have two factors A and B , we say that the data set is balanced with respect to (joint) factor AB if the number of individuals at each level of AB is the same for all possible levels of A and B .

With data in a classificatory factorial structure, it is natural to index individuals by the levels of the several factors which the individual possesses. So with only one factor, A , we may index individuals by $i(j)$, or simply by ij without causing any confusion, in which i denotes the level of the factor A of the individual and j indexes individuals within the subclass of individuals having level i of A . We may note that $i(j)$ indexes subclasses which each consists of one individual.

4.12.4 Balanced One-Way Classification

We have just one factor of classification, A , and we index the levels of this factor by $i = 1, 2, \dots, A$. Our use of the letter A for two purposes should not cause any confusion. To use different letters is easy but very tedious. We suppose that there are r individuals in each subclass, we denote the totality of observations by

$$\{y_{ij}: i = 1, 2, \dots, A, j = 1, 2, \dots, r\}.$$

Now we note that we can form an average within each i class, which we denote by $\bar{y}_{i.}$ and an overall average which we denote by $\bar{y}_{..}$. Then, obviously, we have the identity

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}).$$

Then we can form

$$\sum_{ij} y_{ij}^2 = \sum_{ij} \bar{y}_{..}^2 + \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{ij} (y_{ij} - \bar{y}_{i.})^2 \text{ plus sums of cross products.}$$

But the sums of cross products give zero:

$$\begin{aligned} \sum_{ij} \bar{y}_{..}(\bar{y}_{i.} - \bar{y}_{..}) &= \bar{y}_{..} \sum_j \sum_i (\bar{y}_{i.} - \bar{y}_{..}) = 0 \\ \sum_{ij} \bar{y}_{..}(y_{ij} - \bar{y}_{i.}) &= \sum_i \bar{y}_{..} \sum_j (y_{ij} - \bar{y}_{i.}) = 0 \\ \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) &= \sum_i (\bar{y}_{i.} - \bar{y}_{..}) \sum_j (y_{ij} - \bar{y}_{i.}) = 0. \end{aligned}$$

**Table 4.5 ANOVA for One-Way Classification
with Equal Numbers**

Source	d.f.	SS
CF	1	$rA\bar{y}_{..}^2$
A	$A - 1$	$r\sum_i(\bar{y}_{i.} - \bar{y}_{..})^2$
Residual	$A(r - 1)$	$\sum_{ij}(y_{ij} - \bar{y}_{i.})^2$
Total	rA	$\sum_{ij}y_{ij}^2$

We have further

$$\sum_{ij}\bar{y}_{..}^2 = rA\bar{y}_{..}^2 = \text{CF},$$

where CF is referred to as the correction factor, and

$$\begin{aligned}\sum_{ij}(\bar{y}_{i.} - \bar{y}_{..})^2 &= r \sum_i(\bar{y}_{i.} - \bar{y}_{..})^2 = \text{SS}(A) \\ \sum_{ij}(y_{ij} - \bar{y}_{i.})^2 &= \text{SS}(\text{Units within } A)\end{aligned}$$

thus giving the ANOVA of Table 4.5. The d.f. can be obtained as follows:

- (i) CF is the square of one linear function so CF has 1 degree of freedom.
- (ii) $r\sum_i(\bar{y}_{i.} - \bar{y}_{..})^2$ is the sum of squares of A linear functions $\sqrt{r}(\bar{y}_{i.} - \bar{y}_{..})$, $i = 1, \dots, A$, which are connected by one linear relation $\sum_i \sqrt{r}(\bar{y}_{i.} - \bar{y}_{..}) = 0$. So this sum of squares is said to have $(A - 1)$ degrees of freedom.
- (iii) Finally, $\sum_{ij}(y_{ij} - \bar{y}_{i.})^2$ is the sum of squares of Ar linear functions that are connected by A linear relations, $\sum_j(y_{ij} - \bar{y}_{i.}) = 0$, $i = 1, \dots, A$ and hence is said to have $Ar - A = A(r - 1)$ degrees of freedom.

Usually the term CF is subtracted from Total without renaming so that

$$\text{new Total} = \sum_{ij}y_{ij}^2 - rA\bar{y}_{..}^2 \quad \text{with d.f.} = rA - 1.$$

4.12.5 Two-Way Classification with Equal Numbers

We have a data set classified by rows (R) and columns (C) with $n(\geq 1)$ observations within each row-column cell. The observations are indexed by $i = 1, \dots, R$; $j = 1, \dots, C$; and $k = 1, \dots, n$, that is, y_{ijk} . We can now have the identity:

$$\begin{aligned}y_{ijk} &= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) \\ &\quad + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.}),\end{aligned}\tag{4.49}$$

in which

$$\bar{y}_{...} = \frac{1}{RCn} \sum_{ijk} y_{ijk}, \quad \bar{y}_{i..} = \frac{1}{Cn} \sum_{jk} y_{ijk}$$

and so on.

Then just as in the simple case in Section 4.12.4 we obtain by squaring both sides of (4.49) and summing over all indices

$$\begin{aligned} \sum_{ijk} y_{ijk}^2 &= \sum_{ijk} \bar{y}_{...}^2 + \sum_{ijk} (\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{ijk} (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &\quad + \sum_{ijk} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{ijk} (y_{ijk} - \bar{y}_{ij.})^2. \end{aligned} \quad (4.50)$$

The cross product terms in the square of the sum yield zero; for example,

$$\begin{aligned} &\sum_{ijk} (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) \\ &= \sum_k \sum_i (\bar{y}_{i..} - \bar{y}_{...}) \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) = 0 \end{aligned}$$

because by the “equal numbers” $\sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) = 0$.

The constituent terms on the RHS of (4.50) are in order: CF, SS(Rows), SS(Cols), SS(Rows \times Cols), and SS(within cells). We thus have the ANOVA as given in Table 4.6.

4.12.6 Experimental versus Observational Studies

We shall digress here briefly from our general development of classificatory structures and elucidate how these structures can occur in different contexts. We shall demonstrate later (Chapter 9) that this has certain consequences regarding inference obtained from the models of such structures. To keep the discussion focused on the essential idea we confine ourselves here to the two-way classification and use the following examples.

EXAMPLE 4.3: An experiment was conducted to determine the effects of three different fungicides (A_1, A_2, A_3) on the yield of fruit from four different cultivars of apple trees (B_1, B_2, B_3, B_4). In an orchard six trees from each cultivar were available for the experiment. Each fungicide was applied, by random assignment, to two trees from each cultivar. Yields of fruit were obtained at the end of the growing season. \square

EXAMPLE 4.4: An experiment was conducted to determine the effect of ozonization at three reaction times (A_1, A_2, A_3) and four pH levels (B_1, B_2, B_3, B_4) on effluent decline. Each combination (A_i, B_j) ($i = 1, 2, 3; j = 1, 2, 3, 4$) was applied to two samples of effluent. \square

Table 4.6 ANOVA for Two-Way Classification with Equal Numbers

Source	d.f.	SS
CF	1	$RCn\bar{y}_{...}^2$
Rows	$R - 1$	$Cn\sum_i(\bar{y}_{i..} - \bar{y}_{...})^2$
Columns	$C - 1$	$Rn\sum_j(\bar{y}_{.j.} - \bar{y}_{...})^2$
Rows \times Columns	$(R - 1)(C - 1)$	$n\sum_{ij}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$
Within Cells	$RC(n - 1)$	$\sum_{ijk}(y_{ijk} - \bar{y}_{ij.})^2$
Total	RCn	$\sum_{ijk}y_{ijk}^2$

EXAMPLE 4.5: A study was conducted to investigate possible differences in milk butter fat for cows in three age groups (A_1, A_2, A_3) from four breeds (B_1, B_2, B_3, B_4). For each combination (A_i, B_j) ($i = 1, 2, 3; j = 1, 2, 3, 4$) two cows were randomly selected and their butterfat percentages determined. \square

The common feature of these three examples is that the observations y_{ijk} ($i = 1, 2, 3; j = 1, 2, 3, 4; k = 1, 2$) are typically displayed in a two-way table as given below:

	B_1	B_2	B_3	B_4
A_1	x, x	x, x	x, x	x, x
A_2	x, x	x, x	x, x	x, x
A_3	x, x	x, x	x, x	x, x

where the x 's represent the two observations for each cell, that is, each combination (A_i, B_j). Furthermore, each y_{ijk} can be expressed in terms of a two-way classification model as

$$y_{ijk} \doteq \mu + a_i + b_j + (ab)_{ij}$$

or, mimicking (4.49),

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk} \quad (4.51)$$

or in terms of a cell means model as

$$y_{ijk} \doteq \mu_{ij}$$

or

$$y_{ijk} = \mu_{ij} + e_{ijk}, \quad (4.52)$$

where a_i represents the effect of A_i ($i = 1, 2, 3$), b_j represents the effect of B_j ($j = 1, 2, 3, 4$), and $(ab)_{ij}$ represents the interaction between A_i and B_j (expressing, for example, differences among the effects of factor A depending on the levels of factor B).

The reader will have noticed that although there appears to be only one linear model, the situations leading to this model are quite different. More specifically, Example 4.3 represents a generalized randomized block design (Section 9.7), where the fungicides represent the levels of the treatment factor, and the cultivars represent the levels of the blocking (or intrinsic) factor. Example 4.4 describes a completely randomized design (Chapter 6) with a factorial treatment structure (Section 11.2), where reaction time and pH level represent the treatment factors. Thus Examples 4.3 and 4.4 describe intervention or experimental studies. On the other hand, Example 4.5 represents an observational study, and only data from such a study should be referred to, strictly speaking, as a "two-way classification".

One may ask why we draw this distinction. The reason will become clear when we discuss the individual experimental designs and how statistical inferences can be drawn from such experiments (for Example 4.5 we do this in Section 4.17). We shall show that the distinction arises because of different properties for the error term e_{ijk} , in model (4.51) as a consequence of the randomization of the treatments to experimental units. For Example 4.3 this will lead to an asymmetry for the treatment effects, a_i , and the block effects, b_j , a crucial distinction arising only in experimental but not in observational studies. This is another reason why we discourage the use of the cell means model like (4.52) for experimental studies since such a model implies symmetry of the various factors.

4.12.7 General Classificatory Structure

To develop a complete picture of the nesting and hence of the crossing relationship, the following is useful. Let the units or a set of individuals be indexed by a variable u . Let A be a factor. Then individual u lies in some level of A . Denote that level by $x_A(u)$. Take a subset S consisting of k factors. Then the joint level of u with respect to S will be a k -vector, $S(u)$. Individuals u and u' will be at the same level of S if $S(u) = S(u')$. Let T be another subset of m factors. Then the level of u with respect to T will be an m -vector, $T(u)$. If whenever $S(u) = S(u')$, it is the case that $T(u) = T(u')$, we say that factor subset S is nested by factor subset T or equivalently that T nests S . It is useful also to define what will be meant by the product of two sets of factors S and T . We define ST to be the set of factors that contains the factors in S and/or in T . So that if $S = ADE$ and $T = ABF$, $ST = ABDEF$. Let S and T be two sets of factors with associated partitions. Then S and T have the same associated partitions if $S(u) = S(u')$ implies $T(u) = T(u')$ and $T(u) = T(u')$ implies $S(u) = S(u')$.

Suppose there are factors A , B and C . Then the totality of formal possible product factors are A , B , AB , C , AC , BC , and ABC . These will be different generalized factors only if there are no nestings. Suppose B is nested by A . Then B gives the same partition as does AB , and BC gives the same partitioning as ABC . The totality of distinct partitionings or factors is then

$$A, AB, C, AC, \text{ and } ABC.$$

It is useful to name a partition by that name with the most letters.

This leads to a useful indexing of levels of factors. If A nests B and if we index the levels of A by i , then we index the levels of B by ij . Then we index units in AB

subclasses by ij , the first index giving the level of A and the second index j indexing levels of B within A levels. And similarly for generalized factors.

These ideas lead to the concept of *admissible means*. A mean is admissible if it is a mean of a subclass resulting from a particular partition. If, for instance, A nests B and B nests C with observations indexed by i = level of A , ij = level of B and ijk = level of C , then the admissible means are $\bar{y}_{...}$, $\bar{y}_{i..}$, $\bar{y}_{ij.}$, and y_{ijk} . If we have two factors A and B that are crossed, with factor C nested by AB combinations, and levels indexed by i, j, k , respectively, the admissible means are $\bar{y}_{...}$, $\bar{y}_{i..}$, $\bar{y}_{.j.}$, $\bar{y}_{ij.}$, and y_{ijk} .

We can transfer the idea of nesting of factors to the idea of nesting of subscripts. So, for example, if A nests B and B nests C and we index by ijk , we say that i nests j , and j nests k .

A useful idea given by Zyskind (1962) is that of the rightmost bracket of a set or a subset of subscripts. Suppose the full set of subscripts is i^1, i^2, \dots, i^m . Then an admissible mean is defined as one in which whenever a nested index occurs, then all the subscripts that nest it must also appear. Let j^1, j^2, \dots, j^r be a subset of the subscripts. Then the group of subscripts (in that subset) that nest no other subscripts is said to be the *rightmost bracket* of the subset. It is convenient to enclose the rightmost bracket of a subset of subscripts in parentheses. If, for example, we have a structure in which A nests B and C is crossed with A and B , with subscripts i, j, k , respectively, the admissible means are denoted by $\bar{y}_{...}$, $\bar{y}_{i..}$, $\bar{y}_{i(j.)}$, $\bar{y}_{..k}$, $\bar{y}_{i.k}$, and $y_{i(jk)}$. Alternatively, we say that a subscript belongs to the rightmost bracket of a group of subscripts if no subscript of the group is nested in it.

We can now give the ANOVA of a *balanced data structure*, which is one such that when cross labeling is used for every factor, whether crossed or nested, the range of any subscript is the same for all possible combinations of values of all the other subscripts. Then with $ij(uv)$, for example, we define the *component associated with this group of subscripts* to be

$$y_{ij(uv)} - \bar{y}_{ij(u.)} - \bar{y}_{ij(.v)} + \bar{y}_{ij..}$$

The sum of such components with fixed ij , and summing over u and/or v is zero because of the balance in the balanced data structure. The idea is that a *component* starts with an admissible mean and contains admissible means given by averaging over subscripts in the rightmost bracket with sign equal to -1 if averaging is over an odd number of subscripts and $+1$ if averaging is over an even number of subscripts.

For a balanced data structure, we can write an identity such as (see Section 4.12.4)

$$y_{i(j)} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{i(j)} - \bar{y}_{i.})$$

and clearly $\sum_i (\bar{y}_{i.} - \bar{y}_{..}) = 0$, $\sum_j (y_{i(j)} - \bar{y}_{i.}) = 0$ for each i . Also we have a sort of orthogonality in that we see that $\sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})(y_{i(j)} - \bar{y}_{i.})$ is zero by summing first on j .

Hence

$$\sum_{ij} y_{i(j)}^2 = \sum_{ij} \bar{y}_{..}^2 + \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{ij} (y_{i(j)} - \bar{y}_{i.})^2.$$

This expresses the total sum of squares of the data in the data structure as

$$\begin{aligned} \text{Total sum of squares} &= \text{correction factor} \\ &+ \text{sum of squares between classes} \\ &+ \text{sum of squares within classes.} \end{aligned}$$

We have a further property exemplified by the following:

$$\begin{aligned} \sum_{ij} (\bar{y}_{i.} - \bar{y}_{..})^2 &= \sum_{ij} \bar{y}_{i.}^2 - \sum_{ij} \bar{y}_{..}^2 \\ \sum_{ij} (y_{i(j)} - \bar{y}_{i.})^2 &= \sum_{ij} y_{i(j)}^2 - \sum_{ij} \bar{y}_{i.}^2 \end{aligned}$$

Identities for balanced data structures contain components associated with each admissible mean, each component being given by averaging over subscripts in the rightmost bracket with the sign chosen as explained above. Each component gives rise to a sum of squares by squaring the component and summing over all subscripts for a single observation. The totality of those sums of squares then constitutes the ANOVA. As an example, consider the structure in which A nests B and C is crossed with A and B . Following Zyskind (1962) we express this data structure symbolically as $(A : B)(C)$, where $:$ denotes the nesting relationship and $(\cdot)(\cdot)$ indicates that factors in different parentheses are crossed. Six admissible means, the associated components and sums of squares are given in Table 4.7.

Table 4.7 Admissible Means, Components, and Sums of Squares for Model $(A : B)(C)$

Means	Components	Sums of Squares
$\bar{y}_{...}$	$\bar{y}_{...}$	$\sum_{ijk} \bar{y}_{...}^2$
$\bar{y}_{i..}$	$\bar{y}_{i..} - \bar{y}_{...}$	$\sum_{ijk} (\bar{y}_{i..} - \bar{y}_{...})^2$
$\bar{y}_{i(j).}$	$\bar{y}_{i(j).} - \bar{y}_{i..}$	$\sum_{ijk} (\bar{y}_{i(j).} - \bar{y}_{i..})^2$
$\bar{y}_{..k}$	$\bar{y}_{..k} - \bar{y}_{...}$	$\sum_{ijk} (\bar{y}_{..k} - \bar{y}_{...})^2$
$\bar{y}_{i.k}$	$\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...}$	$\sum_{ijk} (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$
$y_{i(jk)}$	$y_{i(jk)} - \bar{y}_{i(j).} - \bar{y}_{i.k} + \bar{y}_{i..}$	$\sum_{ijk} (y_{i(jk)} - \bar{y}_{i(j).} - \bar{y}_{i.k} + \bar{y}_{i..})^2$
	$y_{i(jk)}$	$\sum_{ijk} y_{i(jk)}^2$

We note that the sum of the components is equal to the individual observation. Hence

$$y_{i(jk)} = \text{sum of components}$$

represents an identity which gives rise to an appropriate linear model for the given data structure. As a consequence of this representation and the assumed balancedness of the

data structure we also have that

$$\sum_{ijk} y_{i(jk)}^2 = \text{sum of sums of squares}$$

as exhibited in Table 4.7.

To help understand data structures such as the one discussed above and others it is useful to use a diagrammatic representation as developed by Throckmorton (1961) (see also Kempthorne et al., 1961). To illustrate this we show the following examples in Figure 4.1:

- (i) A (one-way classification),
- (ii) $A : B$ (two-fold nested classification, that is, B nested in A),
- (iii) $(A)(B)$ (two-way crossed classification),
- (iv) $[(A)(B)] : C$ (A and B crossed, C nested within AB),
- (v) $(A : B)(C)$ (B nested in A and C crossed with A and B).

In these *structure diagrams* μ indicates the overall population to be partitioned according to the structure and ε indicates the fully indexed individual observation.

4.12.8 The Well-Formulated Model

In order to fully understand the concepts of crossed and nested factors (see Section 4.12.6) and to understand the impact of these notions on the formulation of appropriate linear models it is useful and important to present the idea of a *well-formulated model*. To do this in complete generality would be rather cumbersome, so we shall illustrate this idea in terms of examples.

Suppose we have factors A, B, C, D, E where factors A, B, C are crossed. D is nested in ABC combinations, and E is nested in BC combinations. The structure diagram is given in Figure 4.2. Then the list of generalized factors is

$$A, B, AB, C, AC, BC, ABC, ABCD, BCE, ABCE, ABCDE.$$

We index combinations by i, j, k, l, m , for levels of A, B, C, D, E and n for the individual in $ABCDE$ subclasses. The admissible means are then

$$\begin{aligned} &\bar{y}_{(i)....}, \quad \bar{y}_{.(j)....}, \quad \bar{y}_{(ij)....}, \quad \bar{y}_{..(k)...} \\ &\bar{y}_{(i).(k)...}, \quad \bar{y}_{.(jk)...}, \quad \bar{y}_{(ijk)...}, \quad \bar{y}_{ijk(l)...}, \\ &\bar{y}_{.jk.(m).}, \quad \bar{y}_{(i)jk.(m).}, \quad \bar{y}_{ijk(lm)}. \end{aligned}$$

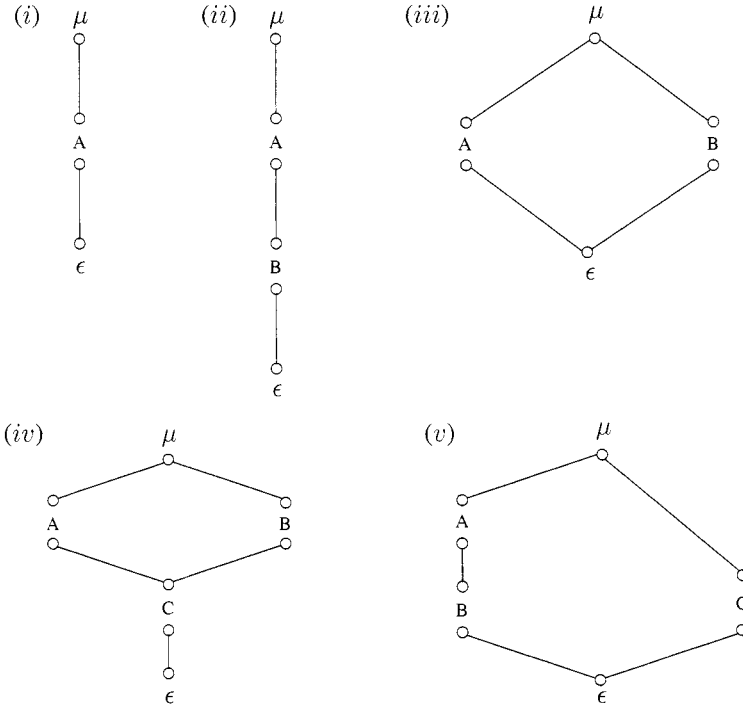


Figure 4.1 Structure diagrams.

For each mean we also give the rightmost bracket (note that some means contain two brackets; this is simply done for convenience to retain the order of the subscripts; in this case the combined brackets constitute the rightmost bracket).

This leads to a scalar model of the following form:

$$y_{ijklm} \doteq \mu + a_i + b_j + (ab)_{ij} + c_k + (ac)_{ik} + (bc)_{jk} \\ + (abc)_{ijk} + (abcd)_{ijkl} + (bce)_{jkm} + (abce)_{ijkm} + (abcde)_{ijklm}. \quad (4.53)$$

Model (4.53) may be called a *full model*, with a term for every subset of every partition. We have to consider models that arise by elimination of factors and combinations of factors. It may be that we should not include a term such as $(ac)_{ik}$. A simple idea is merely to remove the terms $(ac)_{ik}$. But if we remove the term $(ac)_{ik}$, we are removing the partition AC . We then note that doing this leaves in the model the term $(abc)_{ijk}$ which comes from the partition ABC . However, the partition ABC is nested by the partition AC , so if the partition AC is to be not included, then so must the partition ABC . We make the definition:

A model is *well-formulated* if whenever a term corresponding to a partition π is included, the terms associated with partitions that nest π are included.

So for instance, with crossed factors A and B , the model

$$y_{ijk} \doteq \mu + a_i + (ab)_{ij}$$

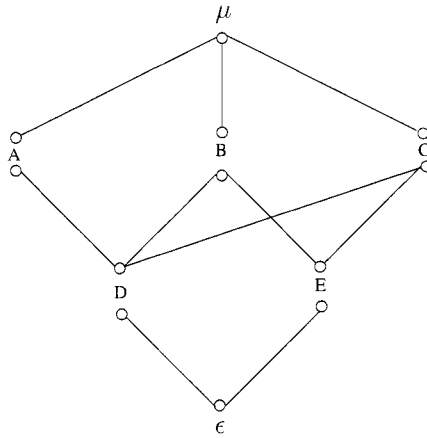


Figure 4.2 Structure diagram for model (4.53).

is not well-formulated. If, however, factor B is nested by factor A , this model is well-formulated.

We insert that discussion of the type above should make it clear why we think that using the means model (see Section 4.11.2) is not appropriate here and in general for more complex structures. The numbers of possible models can be substantial and our arguments spell out clearly how the various models can be derived. Moreover, in Chapters 9, 10, and 13, we shall show that in the context of intervention studies the factors for a given data structure are not always symmetric with respect to statistical inference. To be more specific, we show, for example, that even though the observations from a randomized complete block design (see Chapter 9) have the structure as given in Figure 4.1 (iii) the factors A (treatments) and B (blocks) have to be treated differently due to the randomization procedure (see also Section 4.12.6).

The normal equations for any well-formulated model and balanced data structure can be written down easily because they take the form

$$\text{Model value of RHS} = \text{Admissible sum}$$

for all admissible sums, corresponding to admissible means listed above, for example for model (4.53). To obtain an analysis of variance, we have to fit the successive models obtained by keeping terms providing that if a term is kept, the terms that nest that term must be kept, also.

Models obtained in this way are not of full rank or, in other words, are over-parametrized. They can be made of full rank by adjoining appropriate conditions on parameters. The obvious choice of conditions is obtained by retaining rightmost brack-

ets. In the above somewhat complicated data structure (4.53), they are as follows:

$$\begin{aligned}
\sum_i a_i &= 0, \sum_j b_j = 0, \sum_k c_k = 0, \\
\sum_i (ab)_{ij} &= 0, \sum_j (ab)_{ij} = 0, \sum_i (ac)_{ik} = 0, \sum_k (ac)_{ik} = 0, \\
\sum_j (bc)_{jk} &= 0, \sum_k (bc)_{jk} = 0, \\
\sum_i (abc)_{ijk} &= 0, \sum_j (abc)_{ijk} = 0, \sum_k (abc)_{ijk} = 0, \\
\sum_l (abcd)_{ijkl} &= 0, \sum_m (bce)_{jkm} = 0 \\
\sum_i (abce)_{ijk m} &= 0, \sum_m (abce)_{ijk m} = 0, \sum_l (abcde)_{ijklm} = 0, \sum_m (abcde)_{ijklm} = 0.
\end{aligned}$$

Note that, as indicated above, the conditions on any parameter involve single summations over all indices in the rightmost bracket for that term (as given for the corresponding admissible mean). Therefore, it is easy to write out these conditions. We shall not pursue this example because to do so would involve a mass of equations which can be solved in reasonable form only if the whole data structure is factor balanced and data balanced.

4.13 UNBALANCED DATA STRUCTURES

We have seen in Section 4.12 that for balanced data structures we can easily derive a well-formulated model, obtain the normal equations and the ANOVA table. Much of the ease of doing this is a consequence of balancedness, in particular solving the NE and then writing out the ANOVA table. In practical situations, in particular for observational studies and to a lesser degree for intervention studies, we do encounter, however, unbalanced data structures. Such structures can lead to certain problems and, in fact, to different approaches for solving the NE and obtaining the ANOVA table. We shall not give a general discussion here, but rather consider some simple structures to illustrate the basic ideas.

4.13.1 Two-Fold Nested Classification

For the two-fold structure, a well-formulated model with natural indexing is

$$y_{ijk} \doteq \mu + a_i + (ab)_{ij}. \quad (4.54)$$

Suppose $i = 1, \dots, A, j = 1, \dots, B_i, k = 1, \dots, n_{ij}$. Then the normal equations are

$$\begin{aligned} n_{..}\mu + \sum_i n_{i.}a_i + \sum_{ij} n_{ij}(ab)_{ij} &= y_{..} \\ n_{i.}\mu + n_{i.}a_i + \sum_j n_{ij}(ab)_{ij} &= y_{i.}, \quad i = 1, \dots, A \\ n_{ij}\mu + n_{ij}a_i + n_{ij}(ab)_{ij} &= y_{ij}, \quad i = 1, \dots, A, \quad j = 1, \dots, B_i \end{aligned}$$

where in an obvious notation $n_{i.} = \sum_j n_{ij}$, $n_{..} = \sum_{ij} n_{ij}$, $y_{ij.} = \sum_k y_{ijk}$, $y_{i.} = \sum_{jk} y_{ijk}$, $y_{..} = \sum_{ijk} y_{ijk}$. These are easy to solve in that we may put $\mu = 0$, $a_i = 0$ ($i = 1, \dots, A$), and a solution is

$$\mu^* = 0, \quad a_i^* = 0, \quad (ab)_{ij}^* = \bar{y}_{ij.},$$

with sum of squares removed equal to $\sum_{ij} \bar{y}_{ij.} y_{ij.}$. Then we have to fit the model in which the $(ab)_{ij}$ terms are deleted and a solution is $\mu^{**} = 0$, $a_i^{**} = y_{i.}/n_{i.}$, with sum of squares removed equal to $\sum_i a_i^{**} y_{i.} = \sum_i y_{i.}^2/n_{i.}$. Finally we have to fit the model: $y_{ijk} \doteq \mu$, with NE solution equal to $\mu^{***} = y_{..}/n_{..}$ and sum of squares removed equal to $y_{..}^2/n_{..}$, the usual correction factor. The ANOVA is then

Source	d.f.	SS
μ	1	$y_{..}^2/n_{..}$
(a_i) after μ	$A - 1$	$\sum_i y_{i.}^2/n_{i.} - y_{..}^2/n_{..}$
$((ab)_{ij})$ after $\mu, (a_i)$	$\sum B_i - A$	$\sum_{ij} y_{ij.}^2/n_{ij} - \sum_i y_{i.}^2/n_{i.}$
Residual	$n_{..} - \sum B_i$	$\sum_{ijk} y_{ijk}^2 - \sum_{ij} y_{ij.}^2/n_{ij}$
Total	$n_{..}$	$\sum_{ijk} y_{ijk}^2$

4.13.2 Two-Way Cross-Classification

We now turn to the two-way data structure. We have two factors A and B such that neither nests the other. A well-formulated model is

$$y_{ijk} \doteq \mu + a_i + b_j + (ab)_{ij}, \quad (4.55)$$

the ranges of the subscripts with such a structure being $i = 1, \dots, A, j = 1, \dots, B, k = 0, 1, 2, \dots, n_{ij}$. We specify that there are n_{ij} data points in cell (ij) , where n_{ij} can be 0. This data structure is very important in the design and analysis of experiments area. A very important design is the incomplete block design in which A is the block factor and B is the treatment factor, and only some treatment levels occur with any particular block level (see Chapter 9). A commonly used model, but not necessarily an appropriate one, is that in which there is no interaction of blocks and treatments, so the terms $(ab)_{ij}$ are zero and the *effect* of a change from level i of A to level i' of A does not vary with the level j of B at which this change takes place. The absence of interaction of this type is critically important in analysis of experiments.

It is important for the reader to understand the difference between models (4.54) and (4.55) as they are associated with different data structures. For example, both models contain the term $(ab)_{ij}$, but they have different meanings. In (4.54) $(ab)_{ij}$ denotes the effect of the j th level of factor B nested in the i th level of factor A , whereas in (4.55) $(ab)_{ij}$ denotes the interaction between the i th level of factor A and the j th level of factor B .

The first important feature of the two-way classification is the matter of identifiability. It is useful here to invoke the cell means model, that is, to define μ_{ij} to be the model value for y_{ijk} . The question then arises with a data structure that has some arbitrary set of cell occupancies, (n_{ij}) , what aspects of the model values are identified. As in all cases of linear models, the question is that of linear identifiability. We may ask then if any particular linear function of $\mu, \{a_i\}, \{b_j\}, \{(ab)_{ij}\}$ is identified. The answer to this question is easily seen by representing a linear function of $\{\mu_{ij}\}$ as $\sum_{ij} c_{ij} \mu_{ij}$ where necessarily summation is over cells (ij) which contain at least one observation. Consider then a linear function of the parameters in model (4.55):

$$d_0\mu + \sum_i d_i a_i + \sum_j f_j b_j + \sum_{ij} h_{ij} (ab)_{ij}.$$

Here the summations extend over all i , all j , and all ij , respectively. This linear function of the parameters is identifiable if there exists a set $\{c_{ij}, ij \text{ occupied}\}$ such that

$$\sum_{ij}^* c_{ij} \mu_{ij} = d_0\mu + \sum_i d_i a_i + \sum_j f_j b_j + \sum_{ij} h_{ij} (ab)_{ij}, \quad (4.56)$$

where Σ^* denotes summation over occupied cells, and the equation holds identically when μ_{ij} is replaced by $\mu + a_i + b_j + (ab)_{ij}$. Comparing the left-hand side and right-hand side of (4.56) we must have

$$\begin{aligned} \sum_{ij}^* c_{ij} &= d_0, & \sum_j^* c_{ij} &= d_i \quad (i = 1, \dots, A) \\ \sum_i^* c_{ij} &= f_j \quad (j = 1, \dots, B), & c_{ij} &= h_{ij} \quad \text{for all cells.} \end{aligned}$$

Further, from these conditions it follows that we must have

$$\sum_i d_i = d_0, \quad \sum_j f_j = d_0, \quad \sum_j h_{ij} = d_i, \quad \sum_i h_{ij} = f_j.$$

Now, in order for μ to be identifiable it follows from (4.56) that we must have $d_0 = 1$, all $d_i = 0$, all $f_j = 0$, and all $h_{ij} = 0$. But because $\sum d_i = d_0$, with $d_0 = 1$, these conditions cannot be satisfied. Hence μ is not identifiable. Also, no linear function of $\{a_i\}$ is identifiable, no linear function of $\{b_j\}$ is identifiable, and a linear function $\sum_{ij} h_{ij} (ab)_{ij}$ is identifiable if and only if h_{ij} is zero for unoccupied cells and $\sum_j^* h_{ij} = 0$, $\sum_i^* h_{ij} = 0$, and $\sum_{ij}^* h_{ij} = 0$. In fact, the only identifiable functions are $\sum_{ij}^* c_{ij} \mu_{ij} = \sum_{ij}^* c_{ij} (\mu + a_i + b_j + (ab)_{ij})$, for any $\{c_{ij}\}$.

For model (4.55) the admissible means are $\bar{y}_{...}, \bar{y}_{i...}, \bar{y}_{.j.}, \bar{y}_{ij.}$. According to the rule given earlier, the NE for this model can be written out easily as follows:

$$\begin{aligned} n_{..}\mu + \sum_i n_{i.}a_i + \sum_j n_{.j}b_j + \sum_{ij} n_{ij}(ab)_{ij} &= y_{...} \\ n_{i.}\mu + n_{i.}a_i + \sum_j n_{ij}b_j + \sum_j n_{ij}(ab)_{ij} &= y_{i..} \quad (i = 1, \dots, A) \\ n_{.j}\mu + \sum_i n_{ij}a_i + n_{.j}b_j + \sum_i n_{ij}(ab)_{ij} &= y_{.j.} \quad (j = 1, \dots, B) \\ n_{ij}\mu + n_{ij}a_i + n_{ij}b_j + n_{ij}(ab)_{ij} &= y_{ij.} \quad (i = 1, \dots, A; j = 1, \dots, B) \end{aligned}$$

Just as for model (4.54) it is easy to solve the NE by putting $\mu = 0, a_i = 0$ ($i = 1, \dots, A$), $b_j = 0$ ($j = 1, \dots, B$). A solution then is

$$\mu^* = 0, \quad a_i^* = 0, \quad b_j^* = 0, \quad (ab)_{ij}^* = \bar{y}_{ij.}.$$

To obtain the ANOVA table we need to fit successively smaller, that is, reduced models. In the context of Section 4.10 model (4.55) is a four-part model. Ordinarily a four-part model would have associated with it $4! = 24$ ordered four-part models and hence 24 associated ANOVAs. This, however, is not the case for the type of classificatory model we are considering here. This is a consequence of the fact that not all such models are well-formulated models. For example, the model

$$y_{ijk} \doteq \mu + a_i + (ab)_{ij}$$

is not a well-formulated model for this case or, for that matter, the *ordered* four-part model

$$y_{ijk} \doteq \mu + a_i + (ab)_{ij} + b_j$$

is not a well-formulated model. In fact, the *only* two well-formulated models are

$$y_{ijk} \doteq \mu + a_i + b_j + (ab)_{ij}$$

and

$$y_{ijk} \doteq \mu + b_j + a_i + (ab)_{ij}.$$

The issues of a well-formulated and overparametrized model point to some difficulties with the derivation of the ANOVA table, or better tables, for the model (4.55) with $n_{ij} \geq 0$ and $n_{ij} = 0$ for some (i, j) . This would require a huge amount of writing. We shall point out, however, that it is in this context that the means model mentioned in Section 4.11.13 is useful. Indeed, it is for situations like this that the means model has received more attention. It allows us to spell out sets of identifiable or estimable functions of the μ_{ij} which may be of interest in explaining data. Since $\hat{\mu}_{ij} = \bar{y}_{ij.}$ and $\text{var}(\hat{\mu}_{ij}) = \sigma^2/n_{ij}$ for $n_{ij} > 0$, it is easy to test hypotheses about $\sum_{ij}^* c_{ij} \mu_{ij}$. We emphasize, however, that it is the fact we are dealing here or in other more complex situations of this sort with the full model, that is, the model that includes all interactions and hence does not require any side conditions, that leads to this ease. Furthermore, we need to distinguish carefully between comparisons (of cell means) for experimental or observation studies (see Section 4.12.6).

4.13.3 Two-Way Classification without Interaction

A special case of model (4.55) occurs when there is no interaction. For such a no-interaction model we can write

$$y_{ijk} \doteq \mu_{ij} = \mu + a_i + b_j. \quad (4.57)$$

Just as with model (4.55) a linear function of the parameters in (4.57) is identifiable if there exists a set $\{c_{ij}, ij \text{ occupied}\}$ such that

$$\sum_{ij}^* c_{ij} \mu_{ij} = d_0 \mu + \sum_i d_i a_i + \sum_j f_j b_j. \quad (4.58)$$

From (4.58) we infer immediately that

$$\sum_{ij}^* c_{ij} = d_0, \quad \sum_j^* c_{ij} = d_i, \quad \sum_i^* c_{ij} = f_j$$

and furthermore

$$\sum_i d_i = d_0, \quad \sum_j f_j = d_0.$$

It is then obvious that, just as for model (4.55), μ in model (4.57) is not identifiable. It is, however, possible now to choose the c_{ij} such that $\sum_{i,j}^* c_{ij} = 0$, not all $d_i = 0$ but, of course, $\sum_i d_i = 0$, and all $f_j = 0$. Thus linear combinations of the a_i , $\sum_i d_i a_i$, are identifiable for different choices of the d_i . These choices depend on the structures of the occupied cells. We shall illustrate this with some examples.

EXAMPLE 4.6: We consider the following 4×4 two-way structure where occupied cells, that is, identifiable μ_{ij} , are marked by x:

		B			
		1	2	3	4
A	1		x	x	
	2	x		x	x
	3	x	x		
	4		x		x

Then, with $c_{21} = 1$, $c_{31} = -1$, all other $c_{ij} = 0$, $a_2 - a_3$ is identified. It is then easy to see, by simply looking at each column in this fashion, that $a_1 - a_3$, $a_1 - a_4$, $a_3 - a_4$, $a_1 - a_2$, and $a_2 - a_4$ are also identified. Obviously, any $a_i - a_{i'}$ ($i \neq i'$; $i, i' = 1, 2, 3, 4$) is identifiable. We say that this data set is *row-connected*. Looking at the rows in the same way we see that $b_1 - b_2$, $b_1 - b_3$, and $b_1 - b_4$, are identified, and hence all $b_j - b_{j'}$ ($j \neq j'$; $j, j' = 1, 2, 3, 4$) are identifiable, a property we refer to as *column-connected*. This exemplifies a rather obvious result: if a two-way data set is row-connected, it is column-connected and vice versa. \square

EXAMPLE 4.7: The simplest example of a connected array is in the diagram:

		B			
		1	2	3	4
A	1	x	x	x	x
	2	x			
	3	x			
	4	x			

□

The point is, simply, that in general, with $y \doteq \mathbf{X}\beta$, if $\mathbf{a}'_1\beta$ and $\mathbf{a}'_2\beta$ are identified, then so are $c_1\mathbf{a}'_1\beta + c_2\mathbf{a}'_2\beta$ for any scalars c_1 and c_2 . This is provable in that if $\mathbf{a}'_1 = \nu'_1\mathbf{X}$, $\mathbf{a}'_2 = \nu'_2\mathbf{X}$, then $c_1\mathbf{a}'_1 + c_2\mathbf{a}'_2 = (c_1\nu'_1 + c_2\nu'_2)\mathbf{X}$. Obviously, the result holds only with respect to rows and columns that are represented in the data set.

EXAMPLE 4.8: It is useful to give a small, nontrivial example of a two-way data set that is not row-connected and equivalently not column-connected:

		B				
		1	2	3	4	5
A	1	x		x		
	2		x	x		
	3				x	x
	4					x
	5	x		x		

We see that columns 1, 2, and 3 are row-connected as are columns 4 and 5, while rows 1, 2, and 5 are column-connected as are rows 3 and 4. This data set consists of two disconnected subsets: rows (1, 2, 5) by columns (1, 2, 3), and rows (3, 4) by columns (4, 5). □

The NE for model (4.57) are easy to write out following the recipe given earlier:

$$\begin{aligned}
 n_{..}\mu + \sum_i n_{i.}a_i + \sum_j n_{.j}b_j &= y_{..} \\
 n_{i.}\mu + n_{i.}a_i + \sum_j n_{ij}b_j &= y_{i.} \quad (i = 1, \dots, A) \\
 n_{.j}\mu + \sum_i n_{ij}a_i + n_{.j}b_j &= y_{.j} \quad (j = 1, \dots, B)
 \end{aligned}$$

We recognize that for this set of $1 + A + B$ equations, summing equations 2 to $(A + 1)$ yields equation 1 and so does summing equations $(A + 2)$ to $(1 + A + B)$. Hence the rank of the coefficient matrix is equal to $A + B - 1$. One way to solve the NE then is to use the method described in Section 4.4.4 [see 2(iii)]: Using the fact that individual a_i 's and b_j 's are not identifiable, we set $a_A = 0$ and $b_B = 0$ and solve the remaining $A + B - 1$ equations in $\mu, a_i (i = 1, \dots, A - 1), b_j (j = 1, \dots, B - 1)$. Following arguments described earlier it is then clear how to proceed further and obtain the ANOVA table or tables for model (4.57) (see also Section 4.7).

4.14 ANALYSIS OF COVARIANCE MODEL

In previous sections we have discussed the linear model $\mathbf{y} \doteq \mathbf{X}\boldsymbol{\beta}$ in general terms. We then focused attention on classificatory models since they play a major role in modeling observations arising from comparative experiments. We now turn to models which incorporate elements from both classificatory and regression models. These models, too, play an important role in the design and analysis of comparative experiments (see Chapter 8). The analysis using such models is generally referred to as *analysis of covariance*.

4.14.1 The Question of Explaining Data

We place this in the context of approximative or explanatory or descriptive linear models. We suppose that we have observations which arise from classificatory data structure with what are called *concomitant variables*. We are interested in the value of classificatory variables and/or concomitant variables, towards explaining or describing the variation in the observation variable.

We can define the questions that are of interest by first writing the linear model

$$\mathbf{y} \doteq \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}, \quad (4.59)$$

where \mathbf{X} represents the values of classificatory variables and \mathbf{Z} represents the concomitant variables. We then partition \mathbf{X} into \mathbf{X}_1 and \mathbf{X}_2 , and \mathbf{Z} into \mathbf{Z}_1 and \mathbf{Z}_2 so that the linear model is

$$\mathbf{y} \doteq \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{Z}_2\boldsymbol{\gamma}_2.$$

Because we are interested in describing or explaining the variation in \mathbf{y} , we would normally include a term $\mathcal{J}\beta_0$ so that a model with useful degree of generality is

$$\mathbf{y} \doteq \mathcal{J}\beta_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{Z}_2\boldsymbol{\gamma}_2.$$

Here is one example (see Chapter 9):

\mathbf{X}_1 is a block incidence matrix,

\mathbf{X}_2 is a treatment incidence matrix,

\mathbf{Z}_1 represents the observations on one concomitant scalar variable,

\mathbf{Z}_2 represents the observations on another concomitant variable.

We are interested, then, in the questions:

- (i) Do $\mathbf{Z}_1, \mathbf{Z}_2$ help explaining the data?
- (ii) Does \mathbf{Z}_1 help?
- (iii) Does \mathbf{Z}_2 help?
- (iv) Do $\mathbf{X}_1, \mathbf{X}_2$ help?

- (v) Does \mathbf{X}_1 help?
- (vi) Does \mathbf{X}_2 help?
- (vii) Do $\mathbf{X}_2, \mathbf{Z}_2$ help?

plus other questions given by altering subscripts of \mathbf{X} s and \mathbf{Z} s.

To answer these questions, we have to ask what they mean. What should we mean by “HELP”? It is obvious that if we wish to describe or explain the variation in \mathbf{y} , we can insert whatever explanatory variables that come to mind. If, for instance, we are trying to describe the variation in a test score over, say, 30 individuals, we might insert the variable SN , the social security number of each individual, and then consider a polynomial in $SN, (SN)^2, \dots, (SN)^{29}$. We would describe the variation in test score perfectly because the residual sum of squares would be zero.

We use least squares fitting to address these questions. As we have seen with $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$ (Section 4.7), we form a quantification of what \mathbf{X}_2 does by considering $SS(\mathbf{X}_2|\mathbf{X}_1)$ and $SS(\mathbf{I}|\mathbf{X}_1\mathbf{X}_2)$. The former measures how much \mathbf{X}_2 helps after \mathbf{X}_1 is used, and the latter measures the residual variation after using \mathbf{X}_1 and \mathbf{X}_2 . So we address the questions by computing the following:

- (i) $SS(\mathbf{Z}_1\mathbf{Z}_2|\mathcal{J} \mathbf{X}_1\mathbf{X}_2)$
- (ii) $SS(\mathbf{Z}_1|\mathcal{J} \mathbf{X}_1\mathbf{X}_2\mathbf{Z}_2)$
- (iii) $SS(\mathbf{Z}_2|\mathcal{J} \mathbf{X}_1\mathbf{X}_2\mathbf{Z}_1)$
- (iv) $SS(\mathbf{X}_1\mathbf{X}_2|\mathcal{J} \mathbf{Z}_1\mathbf{Z}_2)$
- (v) $SS(\mathbf{X}_1|\mathcal{J} \mathbf{X}_2\mathbf{Z}_1\mathbf{Z}_2)$
- (vi) $SS(\mathbf{X}_2|\mathcal{J} \mathbf{X}_1\mathbf{Z}_1\mathbf{Z}_2)$
- (vii) $SS(\mathbf{X}_2\mathbf{Z}_2|\mathcal{J} \mathbf{X}_1\mathbf{Z}_1)$

We may note in passing that there are 14 possible interesting sums of squares with the formulation we are presenting, given by $S(\alpha|\text{rest})$ where α is one of

$$\begin{aligned} &\mathbf{X}_1, \mathbf{X}_2, \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{X}_1\mathbf{X}_2, \mathbf{X}_1\mathbf{Z}_1, \mathbf{X}_1\mathbf{Z}_2, \mathbf{X}_2\mathbf{Z}_1, \mathbf{X}_2\mathbf{Z}_2, \mathbf{Z}_1\mathbf{Z}_2, \\ &\mathbf{X}_1\mathbf{X}_2\mathbf{Z}_1, \mathbf{X}_1\mathbf{X}_2\mathbf{Z}_2, \mathbf{X}_1\mathbf{Z}_1\mathbf{Z}_2, \mathbf{X}_2\mathbf{Z}_1\mathbf{Z}_2 \end{aligned}$$

We compare each of these sum of squares with $SS(\mathbf{I}|\mathcal{J} \mathbf{X}_1\mathbf{X}_2\mathbf{Z}_1\mathbf{Z}_2)$. Just how we can compare these will be discussed in Section 4.17, in which we give an exposition of tests of significance. There is, in fact, simple arithmetic if the factors represented by \mathbf{X}_1 and \mathbf{X}_2 are orthogonal partitions. In that case we have a simple ANOVA that we shall describe now.

4.14.2 Obtaining the ANOVA Table

The general situation is given by the model (4.59):

$$\mathbf{y} \doteq \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}.$$

We suppose that fitting $\mathbf{X}\boldsymbol{\beta}$ and the ANOVA associated with $\mathbf{X}\boldsymbol{\beta}$ can be specified easily. We then know that the NE for (4.59) are

$$\begin{aligned}\mathbf{X}'\mathbf{X}\mathbf{b} + \mathbf{X}'\mathbf{Z}\mathbf{g} &= \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{X}\mathbf{b} + \mathbf{Z}'\mathbf{Z}\mathbf{g} &= \mathbf{Z}'\mathbf{y},\end{aligned}$$

with solution given by

$$\begin{aligned}\mathbf{X}\mathbf{b} &= \mathbf{P}_{\mathbf{X}}(\mathbf{y} - \mathbf{Z}\mathbf{g}) \\ \mathbf{Z}'[\mathbf{I} - \mathbf{P}_{\mathbf{X}}]\mathbf{Z}\mathbf{g} &= \mathbf{Z}'[\mathbf{I} - \mathbf{P}_{\mathbf{X}}]\mathbf{y}.\end{aligned}\tag{4.60}$$

For purposes of identifiability we must have that $\mathbf{Z}'[\mathbf{I} - \mathbf{P}_{\mathbf{X}}]\mathbf{Z}$ is of full rank. The sum of squares of \mathbf{y} resulting from fitting $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$ is equal to

$$\mathbf{y}'\mathbf{P}_{\mathbf{X}}\mathbf{y} + \mathbf{g}'\mathbf{Z}'[\mathbf{I} - \mathbf{P}_{\mathbf{X}}]\mathbf{y}.\tag{4.61}$$

Suppose now that in (4.59) we have $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ and, conformably, $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)$. And we wish to compare the models

$$\mathbf{y} \doteq \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$$

and

$$\mathbf{y} \doteq \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}\boldsymbol{\gamma}.\tag{4.62}$$

To return to our introductory comments, we wish to assess how much \mathbf{X}_2 helps explaining the variability of \mathbf{y} . In typical applications \mathbf{X}_2 represents a treatment incidence matrix and $\boldsymbol{\beta}_2$ represents the vector of treatment effects, $\boldsymbol{\tau}$ (see Chapters 8 and 9). What is needed then is $SS(\mathbf{X}_2|\mathbf{X}_1, \mathbf{Z})$ which can be obtained by using the method explained in Section 4.7. To this end we need to determine the sum of squares from fitting model (4.62), and to determine the sum of squares we have the ANOVA of \mathbf{y} , a sum of squares $\mathbf{y}'\mathbf{P}_{\mathbf{X}_1}\mathbf{y}$, and the RNE for $\boldsymbol{\gamma}$:

$$\mathbf{Z}'[\mathbf{I} - \mathbf{P}_{\mathbf{X}_1}]\mathbf{Z}\tilde{\mathbf{g}} = \mathbf{Z}'[\mathbf{I} - \mathbf{P}_{\mathbf{X}_1}]\mathbf{y}.\tag{4.63}$$

The result then is

$$\mathbf{y}'\mathbf{P}_{\mathbf{X}_1}\mathbf{y} + \tilde{\mathbf{g}}'\mathbf{Z}'[\mathbf{I} - \mathbf{P}_{\mathbf{X}_1}]\mathbf{y}.\tag{4.64}$$

To obtain $SS(\mathbf{X}_2|\mathbf{X}_1, \mathbf{Z})$ we simply take the difference between (4.61) and (4.64).

4.14.3 The Case of One Covariate

In order to illustrate the simple general structure of the arithmetic or algebra of the *analysis of covariance* we consider first the usual structure where \mathbf{Z} is an $n \times 1$ vector \mathbf{z} and γ is a scalar, γ . Using the solutions to the RNE for γ , that is, (4.60) and (4.63), given by

$$g = \frac{\mathbf{z}'[\mathbf{I} - \mathbf{P}_X]\mathbf{y}}{\mathbf{z}'[\mathbf{I} - \mathbf{P}_X]\mathbf{z}}$$

and

$$\tilde{g} = \frac{\mathbf{z}'[\mathbf{I} - \mathbf{P}_{X_1}]\mathbf{y}}{\mathbf{z}'[\mathbf{I} - \mathbf{P}_{X_1}]\mathbf{z}},$$

respectively, we obtain

$$SS(\mathbf{X}_2|\mathbf{X}_1, \mathbf{z}) = \mathbf{y}'(\mathbf{P}_X - \mathbf{P}_{X_1})\mathbf{y} + \frac{[\mathbf{z}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}]^2}{\mathbf{z}'[\mathbf{I} - \mathbf{P}_X]\mathbf{z}} - \frac{[\mathbf{z}'(\mathbf{I} - \mathbf{P}_{X_1})\mathbf{y}]^2}{\mathbf{z}'[\mathbf{I} - \mathbf{P}_{X_1}]\mathbf{z}}. \quad (4.65)$$

By inspection we see that (4.65) contains different types of sums of squares and sums of products in \mathbf{y} and \mathbf{z} . These can be represented as in Table 4.8(a). For brevity, the SS and SP of Table 4.8(a) are renamed in Table 4.8(b). The SS and SP for the model $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{z}\gamma$ are obtained by simply amalgamating the $\mathbf{X}_2|\mathbf{X}_1$ and $\mathbf{I}|\mathbf{X}_1\mathbf{X}_2$ lines in Table 4.8(b) as given in Table 4.8(c). Then (4.65) can be written as

$$SS(\mathbf{X}_2|\mathbf{X}_1, \mathbf{z}) = T_{yy} + E_{yy} - \frac{(T_{yz} + E_{yz})^2}{T_{zz} + E_{zz}} - \left\{ E_{yy} - \frac{E_{yz}^2}{E_{zz}} \right\}. \quad (4.66)$$

To form an opinion on whether $\mathbf{X}_2\beta_2$ is useful after including $\mathbf{X}_1\beta_1 + \mathbf{z}\gamma$ we have to compare (4.66) with

$$E_{yy} - \frac{E_{yz}^2}{E_{zz}},$$

which is the remainder sum of squares for the model $\mathbf{y} = \mathbf{X}\beta + \mathbf{z}\gamma$. We shall discuss how this comparison may be made in Section 4.17 and give more details on this whole procedure in connection with specific experimental designs (see Chapters 8 and 9).

4.14.4 The Case of Several Covariates

To conclude this section we comment briefly on the case of more than one covariate, say m covariates. For this purpose we write $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m)$, $\gamma' = (\gamma_1, \gamma_2, \dots, \gamma_m)$, and $\mathbf{g}' = (g_1, g_2, \dots, g_m)$ in (4.59). Then the RNE for γ as given in (4.60) consists of m scalar equations

$$\begin{pmatrix} \mathbf{z}'_1 \\ \mathbf{z}'_2 \\ \vdots \\ \mathbf{z}'_m \end{pmatrix} [\mathbf{I} - \mathbf{P}_X](\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m) \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{pmatrix} = \begin{pmatrix} \mathbf{z}'_1 \\ \mathbf{z}'_2 \\ \vdots \\ \mathbf{z}'_m \end{pmatrix} [\mathbf{I} - \mathbf{P}_X]\mathbf{y}.$$

Table 4.8 Auxiliary ANOVAs

<i>(a) Sums of Squares and Sums of Products for $y \doteq X_1\beta_1 + X_2\beta_2 + z\gamma$</i>			
Source	SS(y)	SP(y, z)	SS(z)
X_1	$y'P_{X_1}y$	$y'P_{X_1}z$	$z'P_{X_1}z$
$X_2 X_1$	$y'[P_X - P_{X_1}]y$	$y'[P_X - P_{X_1}]z$	$z'[P_X - P_{X_1}]z$
$I X_1X_2$	$y'[I - P_X]y$	$y'[I - P_X]z$	$z'[I - P_X]z$
<i>(b) Symbolic Expressions for SS(y), SP(y, z), and SS(z)</i>			
X_1	A_{yy}	A_{yz}	A_{zz}
$X_2 X_1$	T_{yy}	T_{yz}	T_{zz}
$I X_1X_2$	E_{yy}	E_{yz}	E_{zz}
<i>(c) Symbolic Expressions for SS(y), SP(y, z), and SS(z) for $y \doteq X_1\beta_1 + z\gamma$</i>			
X_1	A_{yy}	A_{yz}	A_{zz}
$I X_1$	$T_{yy} + E_{yy}$	$T_{yz} + E_{yz}$	$T_{zz} + E_{zz}$

The i th equation ($i = 1, 2, \dots, m$) is given by

$$\sum_j z'_i[I - P_X]z_j g_j = z'_i[I - P_X]y.$$

We note that all the arithmetic that is involved here can be represented by the ANOVA of y, z_1, z_2, \dots, z_m and partition of the sum of products of y and each z_i . Furthermore, the partition of the sum of products of y and a particular z , say z_j , is given by the ANOVA of $y + z$, with for any source in the ANOVA

$$SS(y + z_j) = SS(y) + SS(z_j) + 2SP(y, z_j),$$

where $SP(y, z_j)$ is defined by, or given by, this equation. For $m = 2$, this is illustrated in Section 8.7.

4.15 FROM DATA ANALYSIS TO STATISTICAL INFERENCE

So far we have described some procedures for analyzing data, that is, looking at data. In that presentation, we start with the data and we adjoin no assumptions. A standard problem is that we wish to make inferences. In the area of the present book, our aim is to establish laws of uncertainty with regard to treatment effects. We shall discuss this in the chapter on randomization (see Chapter 5). This outlook and procedure stands outside the standard ideas of mathematical statistics. It is based, however, on ideas

coming out of general mathematical statistics. It is necessary therefore, to describe the general approach of that area.

The approach uses the assumption exemplified in the standardly exposted case, that our data x_1, x_2, \dots, x_n is a realization of n random variables X_1, X_2, \dots, X_n that are distributed independently according to $N(\mu, \sigma^2)$ (or whatever). One must ask where this assumption comes from. It permeates statistical theory and practice. It is the basis of all theory whether frequentist or Bayesian. The frequentist approach in general says the data set D , say, is a realization of some random entity that has a distribution F_D which depends on a parameter, scalar or vector, Θ . The Bayesian approach adjoins the assumption that Θ is a realization of some random entity that has a distribution, say G_Θ , which is fixed or depends on a parameter scalar or vector, ψ . And so on, leading to what is called *hierarchical Bayes* process.

What are the problems? The obvious one is the assumption that observations are a realization of independent random variables. This is the initial assumption and it is surely of very doubtful status. To exemplify this we consider as a simple example the agronomic field experiment.

Suppose one wishes to compare the yields of two varieties of, say, corn. One finds, by some process that can be described, a piece of agricultural land. One partitions this land into plots. We need not mention blocking because that idea is irrelevant to the point under discussion. We put variety one on some plots and variety two on other plots. We get observations $\{O_{1j}\}$ and $\{O_{2j}\}$. It is easy, absurdly easy, to say that an appropriate model is

$$O_{ij} = \mu_i + e_{ij}$$

with $\{e_{ij}\}$ being independent realizations of the mathematical random variable X which is distributed as $N(0, \sigma^2)$. With this assumption one can apply the statistical tests, etc., the so-called inference procedures of a first course in statistics. But what is the justification for this assumption? The field plots are not a random sample of *any* population.

It may be intuitively reasonable to take the view that the field plots chosen are the outcome of some stochastic process. In recent years, ideas of spatial statistics and spatial random-processes are being put forward for this experimental situation. The problem that immediately arises is that there is a huge number (indeed, an uncountable infinity) of such processes, each being described by a *sentence*. To use these ideas one has to make a choice of what assumptions to make.

The approach followed by essentially every writer is to make use of the simple normal stochastic linear model.

4.16 SIMPLE NORMAL STOCHASTIC LINEAR MODEL

4.16.1 The Notion of Estimability

We now turn to the basic case of a stochastic linear model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, in which $\mathbf{X}\boldsymbol{\beta}$ is a fixed unknown vector in $C(\mathbf{X})$, \mathbf{e} and hence \mathbf{y} are random vectors in \mathbf{R}^n . Rather

naturally we assume that the expectation of \mathbf{e} , $E(\mathbf{e})$, is null, because with $E(\mathbf{e}) = \boldsymbol{\mu}$, a nonnull vector, we would have

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + [\mathbf{e} - E(\mathbf{e})]$$

or

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{e}},$$

with $E(\tilde{\mathbf{e}}) = \mathbf{0}$ and then $E(\mathbf{y}) = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta}$ is an unknown vector in the linear variate $\{\boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta}\}$. So we consider the stochastic model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $E(\mathbf{e}) = \mathbf{0}$. We then have $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. We suppose that $\boldsymbol{\beta}$ is some unknown fixed vector in \mathbf{R}^p which is not restricted *a priori* in any way. As discussed briefly in Section 14.4, we say that a parametric function, $\boldsymbol{\lambda}'\boldsymbol{\beta}$, is estimable if there exists a vector \mathbf{a} , such that $E(\mathbf{a}'\mathbf{y}) = \boldsymbol{\lambda}'\boldsymbol{\beta}$. Obviously, this tells us that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if and only if there exists an \mathbf{a} such that $\boldsymbol{\lambda}' = \mathbf{a}'\mathbf{X}$, or $\boldsymbol{\lambda} \in R(\mathbf{X})$, recalling that $R(\mathbf{X}) = \{\boldsymbol{\xi}, \boldsymbol{\xi}' = \boldsymbol{\nu}'\mathbf{X} \text{ for some } \boldsymbol{\nu}\}$. This tells us also that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if and only if $\mathbf{X}'(\mathbf{X}')^{-}\boldsymbol{\lambda} = \boldsymbol{\lambda}$.

The general idea of estimability is useful in many contexts. Consider, for instance, the 2-part model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$. We may ask if a parametric function, $\boldsymbol{\lambda}_1'\boldsymbol{\beta}_1$ is estimable. For this to happen we must have that there exists an \mathbf{a}_1 such that

$$\mathbf{a}_1'\mathbf{X}_1 = \boldsymbol{\lambda}_1', \quad \mathbf{a}_1'\mathbf{X}_2 = \mathbf{0}.$$

We can write this in various ways, of which the following is informative. Using M-P inverses we need

$$\mathbf{X}_1'\mathbf{a}_1 = \boldsymbol{\lambda}_1,$$

which implies

$$\mathbf{a}_1 = (\mathbf{X}_1^+)' \boldsymbol{\lambda}_1 + [\mathbf{I} - (\mathbf{X}_1^+)' \mathbf{X}_1'] \boldsymbol{\gamma},$$

for some $\boldsymbol{\gamma}$. So we must have

$$\mathbf{X}_2'(\mathbf{X}_1^+)' \boldsymbol{\lambda}_1 + \mathbf{X}_2'[\mathbf{I} - (\mathbf{X}_1^+)' \mathbf{X}_1'] \boldsymbol{\gamma} = \mathbf{0}$$

or

$$\mathbf{X}_2' = \mathbf{X}_2'(\mathbf{X}_1^+)' \mathbf{X}_1', \quad \mathbf{X}_2'(\mathbf{X}_1^+)' \boldsymbol{\lambda}_1 = \mathbf{0},$$

a potentially useful statement of what must hold for $\boldsymbol{\lambda}_1'\boldsymbol{\beta}_1$ to be estimable.

Rather clearly, we have gone as far as we can go without additional assumptions. The next natural step is to assume that the random vector \mathbf{e} possesses a variance matrix, that is, $E(\mathbf{e}\mathbf{e}')$, an $n \times n$ matrix, the variance matrix of \mathbf{y} , exists.

4.16.2 Gauss-Markov Linear Model

The natural initial assumption to consider is $E(\mathbf{e}\mathbf{e}') = \sigma^2\mathbf{I}$. We give the following definition:

Definition 4.1. The *Gauss-Markov Linear Model* (GLM) is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad E(\mathbf{e}) = \mathbf{0}, \quad E(\mathbf{e}\mathbf{e}') = \sigma^2\mathbf{I},$$

where \mathbf{X} is the known and fixed model matrix and $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector parameter that may take any value in \mathbf{R}^p . \square

The following basic theorem is one of the most important results in linear model theory.

Theorem 4.1 (*The Gauss-Markov Theorem*). With the GLM, the best (minimum variance) linear unbiased estimator (BLUE) $\widehat{\lambda'\beta}$ of an estimable $\lambda'\beta$ is given by $\lambda'b$ where b is any solution of the NE: $X'Xb = X'y$.

Proof. We know $\lambda' = a'X$, $\lambda'\beta = a'X\beta$, $Xb = P_X y$, $\lambda'b = a'P_X y$. Consider the linear estimator $(a'P_X + \delta')y$ for arbitrary δ . For unbiasedness we must have $(a'P_X + \delta')X\beta = \lambda'\beta = a'X\beta$. With β free this tells us that $\delta'X = 0$ or $\delta'Xb = 0$ or $\delta'P_X = 0$. Hence

$$\begin{aligned}\text{var}[(a'P_X + \delta')y] &= \sigma^2(a'P_X + \delta')(P_X a + \delta) \\ &= \sigma^2(a'P_X a + 2\delta'P_X a + \delta'\delta) \\ &= \sigma^2(a'P_X a + \delta'\delta).\end{aligned}$$

This is minimized with respect to δ , obviously, by taking $\delta = 0$. So the result is proved.

The result of Theorem 4.1 implies, of course, that

$$E(\widehat{\lambda'\beta}) = E(\lambda'b) = \lambda'\beta.$$

For later purposes it is useful to obtain also an expression for $\text{var}(\widehat{\lambda'\beta})$. Using results from Sections 4.4.2 and 4.4.4 we find

$$\begin{aligned}\text{var}(\widehat{\lambda'\beta}) &= \text{var}(\lambda'b) \\ &= \text{var}(a'Xb) \\ &= \text{var}[a'X(X'X)^-X'y] \\ &= a'X(X'X)^-X'X(X'X)^{-1}X'a\sigma^2 \\ &= a'P_X P_X' a\sigma^2 \\ &= a'P_X a\sigma^2 \\ &= a'X(X'X)^-X'a\sigma^2 \\ &= \lambda'(X'X)^-\lambda\sigma^2.\end{aligned}$$

In words: the g -inverse $(X'X)^-$ acts as the variance-covariance matrix (apart from σ^2) for estimable functions, independent of which g -inverse is used.

We next give a very important generalization of the Gauss-Markov Theorem.

Theorem 4.2 (*The Aitken Theorem*). Suppose $y = X\beta + e$, $E(e) = 0$, $E(ee') = V\sigma^2$ with V invertible, then the BLUE $\lambda'\beta$ of an estimable $\lambda'\beta$ is $\lambda'b$ where b satisfies the Aitken equation

$$X'V^{-1}Xb = X'V^{-1}y. \quad (4.67)$$

Proof. Since \mathbf{V} is a real symmetric matrix, there exists an orthogonal matrix \mathbf{O} such that

$$\mathbf{O}'\mathbf{V}\mathbf{O} = \mathbf{D} = \begin{pmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix} = \text{diag}(d_1, d_2, \dots, d_n).$$

Then because \mathbf{V} is an invertible variance matrix, the elements $d_i (i = 1, \dots, n)$ are positive and hence have square roots, so we can form $\mathbf{D}^{1/2}$ where we take the positive root always. Then

$$\mathbf{V} = \mathbf{O}\mathbf{D}\mathbf{O}' = \mathbf{O}\mathbf{D}^{1/2}\mathbf{O}'\mathbf{O}\mathbf{D}^{1/2}\mathbf{O}' = \mathbf{V}^{1/2}\mathbf{V}^{1/2} \text{ where } (\mathbf{V}^{1/2})' = \mathbf{V}^{1/2}.$$

Consider then

$$\mathbf{V}^{-1/2}\mathbf{y} = \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-1/2}\mathbf{e},$$

which is a linear model, and can be written (using obvious notation) as

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^*$$

with

$$E(\mathbf{e}^*) = \mathbf{0}, \quad E(\mathbf{e}^*\mathbf{e}^{*\prime}) = \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}\sigma^2 = \mathbf{I}\sigma^2.$$

Hence the derived model is a GLM. In this case $\mathbf{X}'\boldsymbol{\beta}$ is estimable if and only if there exists an \mathbf{a}^* such that $\mathbf{a}^{*\prime}\mathbf{X}^* = \boldsymbol{\lambda}'$ or $(\mathbf{a}^{*\prime}\mathbf{V}^{-1/2})\mathbf{X} = \boldsymbol{\lambda}'$ which is equivalent to $\mathbf{a}'\mathbf{X} = \boldsymbol{\lambda}'$ with $\mathbf{a}' = \mathbf{a}^{*\prime}\mathbf{V}^{-1/2}$. We then apply the Gauss-Markov Theorem and we know that the BLUE of an estimable function $\mathbf{X}'\boldsymbol{\beta}$ is $\mathbf{X}'\mathbf{b}$, where

$$\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{b} = \mathbf{X}^{*\prime}\mathbf{y}$$

or

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

4.16.3 Ordinary Least Squares and Best Linear Unbiased Estimators

An interesting question is: When is the so-called *ordinary least squares* (OLS) *estimator* for an estimable $\mathbf{X}'\boldsymbol{\beta}$, obtained from (4.4), also BLUE, as obtained from (4.67), which is referred to as *generalized least squares* (GLS) *estimator*.

We can address this very easily. If $\mathbf{X}\mathbf{b} = \mathbf{P}_\mathbf{X}\mathbf{y}$ satisfies the Aitken equation (4.67) for all \mathbf{y} , then the following equations result:

$$\begin{aligned} \mathbf{X}'\mathbf{V}^{-1}\mathbf{P}_\mathbf{X} &= \mathbf{X}'\mathbf{V}^{-1} \\ \mathbf{P}_\mathbf{X}\mathbf{V}^{-1}\mathbf{P}_\mathbf{X} &= \mathbf{P}_\mathbf{X}\mathbf{V}^{-1} \\ \mathbf{P}_\mathbf{X}\mathbf{V}^{-1} &= \mathbf{V}^{-1}\mathbf{P}_\mathbf{X} \\ \mathbf{V}\mathbf{P}_\mathbf{X} &= \mathbf{P}_\mathbf{X}\mathbf{V}. \end{aligned}$$

Hence

$$\mathbf{V}\mathbf{X} = \mathbf{X}\mathbf{Q}$$

for some \mathbf{Q} for the least squares estimator to be BLUE. Contrariwise, if $\mathbf{V}\mathbf{X} = \mathbf{X}\mathbf{Q}$ for some \mathbf{Q} , then

$$\mathbf{V}\mathbf{X} = \mathbf{X}\mathbf{Q}$$

implies

$$\mathbf{V}\mathbf{X}\mathbf{B} = \mathbf{X}\mathbf{Q}\mathbf{B}$$

or

$$\mathbf{V}\mathbf{P}_\mathbf{X} = \mathbf{X}\mathbf{Q}\mathbf{B}$$

(where $\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{X}'$) and

$$\mathbf{P}_\mathbf{X}\mathbf{V}\mathbf{P}_\mathbf{X} = \mathbf{V}\mathbf{P}_\mathbf{X}$$

so

$$\mathbf{V}\mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{X}\mathbf{V}$$

and by transposition

$$\begin{aligned}\mathbf{V}^{-1}\mathbf{P}_\mathbf{X} &= \mathbf{P}_\mathbf{X}\mathbf{V}^{-1} \\ \mathbf{X}'\mathbf{V}^{-1}\mathbf{P}_\mathbf{X}\mathbf{y} &= \mathbf{X}'\mathbf{P}_\mathbf{X}\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{B}\mathbf{y}) &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.\end{aligned}$$

So $\mathbf{b} = \mathbf{B}\mathbf{y}$ satisfies the Aitken equation and $\mathbf{X}\mathbf{b} = \mathbf{X}\mathbf{B}\mathbf{y} = \mathbf{P}_\mathbf{X}\mathbf{y}$ satisfies the Aitken equation. Estimable $\mathbf{X}'\boldsymbol{\beta}$ is $\mathbf{a}'\mathbf{X}\boldsymbol{\beta}$ with OLS estimator $\mathbf{a}'\mathbf{P}_\mathbf{X}\mathbf{y}$ which is the same as what the Aitken equation gives.

The result given above is particularly important in the context of linear models for data from intervention studies. We shall show, starting with Chapter 6, that the derived linear models associated with the various error-control designs have a variance-covariance structure of the form $\mathbf{V}\sigma^2$ with $\mathbf{V} \neq \mathbf{I}$, but with \mathbf{V} such that OLS estimators are, indeed, BLUE.

In discussing the Aitken Theorem we have assumed that \mathbf{V} is nonsingular. Without going into any detail we shall mention briefly the case where \mathbf{V} is singular. Obviously, we cannot form the Aitken equation. It can be shown, however, that the BLUE of $\mathbf{X}\boldsymbol{\beta}$ is given by $\mathbf{A}'\mathbf{y}$, where

$$\begin{aligned}\mathbf{V}\mathbf{A} + \mathbf{X}\mathbf{M} &= \mathbf{0} \\ \mathbf{X}'\mathbf{A} &= \mathbf{X}\end{aligned}$$

for some matrix \mathbf{M} . There is a considerable literature on this [Zyskind (1967), Rao (1967, 1971, 1973), Kempthorne (1971, 1972, 1973a,b, 1976), Watson (1967), and Kempthorne and Doerfler (1969)].

We shall merely state, without proof,

- (i) The OLS estimator and the BLUE are identical if and only if $\mathbf{V}\mathbf{X} = \mathbf{X}\mathbf{Q}$, and
- (ii) the class of matrices \mathbf{V} such that the BLUE of an estimable $\mathbf{X}'\boldsymbol{\beta}$ as given by OLS is

$$\mathbf{V} = c_0\mathbf{I} + \mathbf{P}_\mathbf{X}\mathbf{A}\mathbf{P}_\mathbf{X} + (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{B}(\mathbf{I} - \mathbf{P}_\mathbf{X}),$$

where \mathbf{A} and \mathbf{B} are arbitrary matrices such that \mathbf{V} is a variance matrix.

4.16.4 Expectation of Quadratic Forms

We can adjoin a few very useful ideas from our present basis. With the model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $E(\mathbf{e}) = \mathbf{0}$ and $E(\mathbf{e}\mathbf{e}') = \mathbf{V}$, we have for a fixed vector \mathbf{a} , $E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{X}\boldsymbol{\beta}$. Also, we can obtain higher moments of the elements of \mathbf{y} . The simple one is $E(\mathbf{y}'\mathbf{A}\mathbf{y})$, the expectation of a fixed quadratic form. We have

$$\mathbf{y}'\mathbf{A}\mathbf{y} = (\mathbf{X}\boldsymbol{\beta} + \mathbf{e})'\mathbf{A}(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})$$

and with $E(\mathbf{e}) = \mathbf{0}$,

$$\begin{aligned} E(\mathbf{y}'\mathbf{A}\mathbf{y}) &= E(\mathbf{X}\boldsymbol{\beta})'\mathbf{A}(\mathbf{X}\boldsymbol{\beta}) + E(\mathbf{e}'\mathbf{A}\mathbf{e}) \\ &= \boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} + E[\text{trace}(\mathbf{e}'\mathbf{A}\mathbf{e})] \\ &= \boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} + E[\text{trace}(\mathbf{A}\mathbf{e}\mathbf{e}')] \\ &= \boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} + \text{trace}[\mathbf{A}E(\mathbf{e}\mathbf{e}')] \\ &= \boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} + \text{trace}[(\mathbf{A}\mathbf{V})] \end{aligned} \quad (4.68)$$

If now $\mathbf{V} = \sigma^2\mathbf{I}$, (4.68) becomes

$$\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} + \sigma^2 \text{trace } \mathbf{A} \quad (4.69)$$

and if \mathbf{A} is symmetric idempotent, (4.69) equals

$$\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} + \sigma^2 \text{rank } \mathbf{A}.$$

This provides useful simple results on the vector of residuals, defined to be $\mathbf{y} - \text{Fit}(\mathbf{X}\boldsymbol{\beta})$. With least squares fitting this is $\mathbf{y} - \mathbf{P}_\mathbf{X}\mathbf{y}$. This residual vector has variance-covariance matrix equal to

$$E(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{e}\mathbf{e}'(\mathbf{I} - \mathbf{P}_\mathbf{X}) = (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{V}(\mathbf{I} - \mathbf{P}_\mathbf{X}). \quad (4.70)$$

If $\mathbf{V} = \sigma^2\mathbf{I}$, (4.70) is $\sigma^2(\mathbf{I} - \mathbf{P}_\mathbf{X})$. The residual sum of squares is

$$[(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}]'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y} = [(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{e}]'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{e}$$

with expectation equal to

$$E \text{trace}[(\mathbf{I} - \mathbf{P}_\mathbf{X})^2(\mathbf{e}\mathbf{e}')] = \sigma^2 \text{trace}(\mathbf{I} - \mathbf{P}_\mathbf{X}) = \sigma^2(n - \text{rank } \mathbf{X}).$$

4.17 DISTRIBUTION THEORY WITH GMNLM

4.17.1 Distributional Properties of $\widehat{\lambda'\boldsymbol{\beta}}$

By Gauss-Markov Normal Linear Model (GMNLM), we mean the stochastic model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad E(\mathbf{e}) = \mathbf{0}, \quad E(\mathbf{e}\mathbf{e}') = \sigma^2\mathbf{I}$$

and $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$, that is \mathbf{e} and hence \mathbf{y} follows the multivariate (n -variable) normal distribution (MVN). We shall throughout use the notation $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to mean that \mathbf{y} is an $n \times 1$ random vector that has a multivariate normal distribution with mean vector equal to $\boldsymbol{\mu}$ and with variance matrix equal to $\boldsymbol{\Sigma}$.

The distribution theory associated with GMNLM is rather straightforward:

- (i) The estimator $\widehat{\boldsymbol{\lambda}'\beta}$ of an estimable $\boldsymbol{\lambda}'\beta = \mathbf{a}'\mathbf{X}\beta$ is $\mathbf{a}'\mathbf{P}_\mathbf{X}\mathbf{y} = \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\rho}'\mathbf{X}'\mathbf{y}$, where $\boldsymbol{\rho}$ satisfies $\mathbf{X}'\mathbf{X}\boldsymbol{\rho} = \boldsymbol{\lambda}$, and follows $N(\boldsymbol{\lambda}'\beta, \sigma^2\boldsymbol{\rho}'\boldsymbol{\lambda})$. Also $\boldsymbol{\rho}'\boldsymbol{\lambda} = \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}$. The equations $\mathbf{X}'\mathbf{X}\boldsymbol{\rho} = \boldsymbol{\lambda}$ are called the *conjugate NE*.
- (ii) Suppose we have k linearly independent estimable functions $\boldsymbol{\lambda}'_i\beta$, $i = 1, 2, \dots, k$. Th

$$\text{cov}(\boldsymbol{\lambda}'_i\beta, \boldsymbol{\lambda}'_j\beta) = \sigma^2\boldsymbol{\rho}'_i\boldsymbol{\lambda}_j = \sigma^2\boldsymbol{\lambda}'_i\boldsymbol{\rho}_j,$$

where $\mathbf{X}'\mathbf{X}\boldsymbol{\rho}_i = \boldsymbol{\lambda}_i$, $\mathbf{X}'\mathbf{X}\boldsymbol{\rho}_j = \boldsymbol{\lambda}_j$. If we write

$$\mathbf{W} = (\boldsymbol{\rho}'_i\boldsymbol{\lambda}_j)$$

then with $\boldsymbol{\theta}' = (\boldsymbol{\lambda}'_1\beta, \dots, \boldsymbol{\lambda}'_k\beta)$ and corresponding $\hat{\boldsymbol{\theta}}$, we have that

$$\hat{\boldsymbol{\theta}} \sim N_k(\boldsymbol{\theta}, \sigma^2\mathbf{W}).$$

Additionally,

$$\frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{W}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\sigma^2} \sim \chi_k^2,$$

where χ_k^2 is a random variable having the chi-squared distribution with k degrees of freedom.

- (iii) The estimator of any estimable function is $\boldsymbol{\rho}'\mathbf{X}'\mathbf{y}$ for some $\boldsymbol{\rho}$. The residual vector is $(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}$. Then any set of linear functions $\{\boldsymbol{\rho}'_i\mathbf{X}'\mathbf{y}\}$ and any set of linear functions $\{\boldsymbol{\nu}'_j(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}\}$ being linear functions of multivariate normal random variables, have a joint multivariate normal distribution which is specified entirely by a mean vector and a variance matrix. But

$$\begin{aligned} \text{cov}(\boldsymbol{\rho}'\mathbf{X}'\mathbf{y}, \boldsymbol{\nu}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}) &= E[\boldsymbol{\rho}'\mathbf{X}'\mathbf{e}\boldsymbol{\nu}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{e}] \\ &= E[\boldsymbol{\rho}'\mathbf{X}'\mathbf{e}\mathbf{e}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\boldsymbol{\nu}] \\ &= \sigma^2\boldsymbol{\rho}'\mathbf{X}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\boldsymbol{\nu} = 0 \end{aligned}$$

because $\mathbf{P}_\mathbf{X}\mathbf{X} = \mathbf{X}$, $\mathbf{X}' = \mathbf{X}'\mathbf{P}_\mathbf{X}$. So any linear function $\boldsymbol{\rho}'\mathbf{X}'\mathbf{y}$ and any linear function $\boldsymbol{\nu}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}$ are uncorrelated and hence independent.

It is then natural to use the following terminology:

The *estimation space* is the set of linear functions $\{\boldsymbol{\rho}'\mathbf{X}'\mathbf{y} : \boldsymbol{\rho} \in R^p\}$, and

the *error space* is the set of linear functions $\{\boldsymbol{\nu}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y} : \boldsymbol{\nu} \in R^n\}$.

Then any function defined on the estimation space is independent of any function defined on the error space. So the estimators of any set of estimable functions is independent of $(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y} = (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{e}$ and hence independent of $\mathbf{y}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}$ which is the residual sum of squares.

4.17.2 Distribution of Sums of Squares

We still have to think about the distribution of the residual sum of squares and of any sum of squares in an ANOVA and the joint distribution of linear and quadratic functions. We can accomplish this task rather easily. Every sum of squares in our ANOVA is of the form $\mathbf{y}'\mathbf{S}_i\mathbf{y}$, where $\mathbf{S}'_i = \mathbf{S}_i = \mathbf{S}_i^2$ is s.i.p. with roots that are 1 or 0. More specifically, with $s = k + 1$ sources, we have, using the notation for the projection matrices in Section 4.11.1,

$$\mathbf{S}_1 = \mathbf{P}_1, \mathbf{S}_i = \mathbf{P}_{12\dots i-1}, (i = 2, \dots, k), \mathbf{S}_s = \mathbf{I} - \mathbf{P}_{12\dots k},$$

with

$$\mathbf{I} = \mathbf{S}_1 + \mathbf{S}_2 + \dots + \mathbf{S}_s$$

and

$$\mathbf{S}_i\mathbf{S}_j = \mathbf{O} \quad (i \neq j).$$

Hence by the standard theorem that with $\{\mathbf{S}_i, \mathbf{S}_i\mathbf{S}_j = \mathbf{S}_j\mathbf{S}_i\}$ there exists a single orthogonal matrix that diagonalizes every \mathbf{S}_i , with s sources in the ANOVA there exists a single $n \times n$ orthogonal matrix

$$\mathbf{O} = (\mathbf{O}_1 : \mathbf{O}_2 : \dots : \mathbf{O}_s),$$

where \mathbf{O}_i is $n \times r_i$, $r_i = \text{rank}(\mathbf{S}_i)$, such that

$$\mathbf{O}'_i\mathbf{S}_i\mathbf{O}_i = \mathbf{I}_{r_i}, \quad \mathbf{O}'_i\mathbf{S}_i\mathbf{O}_j = \mathbf{O}, \quad i \neq j. \quad (4.71)$$

(see e.g. Harville, 1997, Section 21.13). The ANOVA comes from

$$\mathbf{y} = \mathbf{S}_1\mathbf{y} + \mathbf{S}_2\mathbf{y} + \dots + \mathbf{S}_s\mathbf{y} \quad (4.72)$$

and then, using the fact that $\mathbf{O}\mathbf{O}' = \mathbf{I}_n$,

$$\mathbf{O}'\mathbf{y} = (\mathbf{O}'\mathbf{S}_1\mathbf{O})\mathbf{O}'\mathbf{y} + (\mathbf{O}'\mathbf{S}_2\mathbf{O})\mathbf{O}'\mathbf{y} + \dots + (\mathbf{O}'\mathbf{S}_s\mathbf{O})\mathbf{O}'\mathbf{y}. \quad (4.73)$$

Because of 4.71 we have

$$\sum_{i=1}^s (\mathbf{O}'\mathbf{S}_i\mathbf{O}) = \text{diag}[\mathbf{O}'_1\mathbf{O}_1, \mathbf{O}'_2\mathbf{O}_2, \dots, \mathbf{O}'_s\mathbf{O}_s],$$

where $\text{diag}[\cdot]$ is a block diagonal matrix. Together with (4.73) this implies that $\mathbf{O}'_i\mathbf{O}_i = \mathbf{I}_{r_i}$. Hence with $\mathbf{z} = \mathbf{O}'\mathbf{y}$, $\mathbf{z}_i = \mathbf{O}'_i\mathbf{y}$, we have, using (4.72) and (4.73), $\mathbf{y}'\mathbf{S}_i\mathbf{y} = (\mathbf{O}'_i\mathbf{y})'(\mathbf{O}'_i\mathbf{y}) = \mathbf{z}'_i\mathbf{z}_i$. Furthermore, $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_s$ are vectors that are distributed according to the multivariate normal distribution. Each will have a certain mean vector; the variance of each vector is $\sigma^2\mathbf{I}$ of appropriate dimensions and the covariance matrix of different vectors is null. Hence the $\{\mathbf{z}_i\}$ are independent. Clearly, then, $\{\mathbf{y}'\mathbf{S}_i\mathbf{y} = \mathbf{z}'_i\mathbf{z}_i\}$ are independent. Finally, we know that if a vector \mathbf{z} is $N(\boldsymbol{\nu}, \mathbf{V})$, that is, is multivariate normal with mean vector $\boldsymbol{\nu}$ and variance matrix \mathbf{V} , then

- (i) $(\mathbf{z} - \boldsymbol{\nu})' \mathbf{V}^{-1} (\mathbf{z} - \boldsymbol{\nu})$ is distributed as χ^2 with degrees of freedom equal to the dimensionality of \mathbf{z} ,
- (ii) $\mathbf{z}' \mathbf{V}^{-1} \mathbf{z}$ is distributed as the noncentral chi-squared distribution with the same number of degrees of freedom and noncentrality parameter $\boldsymbol{\nu}' \mathbf{V}^{-1} \boldsymbol{\nu}$ [or $\frac{1}{2}$ of this with an alternative definition].

From this, we know the distribution under GMNLM of every sum of squares. In our case $\mathbf{V} = \sigma^2 \mathbf{I}$ of appropriate dimension. Hence, under GMNLM with sums of squares, SS_1, SS_2, \dots, SS_s it is the case that SS_i/σ^2 is distributed with its associated degrees of freedom, r_i , according to the noncentral χ^2 distribution. Also the separate such variables are independent. A particular sum of squares has notable behavior, namely, the residual sum of squares with $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$: This is equal to $[(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}]'[(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}]$ but $(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y} = (\mathbf{I} - \mathbf{P}_\mathbf{X})(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{e}$ which has zero expectation. Hence we have, under GMNLM,

$$\frac{\text{SS (residuals)}}{\sigma^2} \sim \chi_{n-r}^2,$$

where $r = \text{rank}(\mathbf{X})$.

4.17.3 Testing of Hypotheses

From the discussion above, we have the following as simple consequences:

- (i) The estimator of any set of estimable functions is distributed independently of the residual sum of squares.
- (ii) The estimator, $\widehat{\boldsymbol{\lambda}'\boldsymbol{\beta}}$, of an estimable function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is such that

$$\frac{\widehat{\boldsymbol{\lambda}'\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\sqrt{\boldsymbol{\rho}'\boldsymbol{\lambda}s^2}} \sim t_{n-r},$$

where, as usual, $\boldsymbol{\rho}$ satisfies $\mathbf{X}'\mathbf{X}\boldsymbol{\rho} = \boldsymbol{\lambda}$, s^2 is [sum of squares of residuals/($n-r$)] and t is a random variable that follows the t -distribution with $(n-r)$ degrees of freedom.

- (iii) $\frac{\widehat{\boldsymbol{\lambda}'\boldsymbol{\beta}}}{\sqrt{\boldsymbol{\rho}'\boldsymbol{\lambda}s^2}} \sim t'_{n-r}$ where t'_{n-r} is a random variable that follows the noncentral t -distribution with noncentrality equal to

$$\frac{\boldsymbol{\lambda}'\boldsymbol{\beta}}{\sqrt{\boldsymbol{\rho}'\boldsymbol{\lambda}\sigma^2}}.$$

- (iv) With $\boldsymbol{\theta}' = (\boldsymbol{\lambda}'_1\boldsymbol{\beta}, \boldsymbol{\lambda}'_2\boldsymbol{\beta}, \dots, \boldsymbol{\lambda}'_m\boldsymbol{\beta})$, a vector of m linearly independent estimable functions, with associated estimator $\widehat{\boldsymbol{\theta}}$ where $\widehat{\boldsymbol{\theta}}' = (\widehat{\boldsymbol{\lambda}'_1\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}'_2\boldsymbol{\beta}}, \dots, \widehat{\boldsymbol{\lambda}'_m\boldsymbol{\beta}})$, and

$$\widehat{\boldsymbol{\lambda}'_i\boldsymbol{\beta}} = \boldsymbol{\rho}'_i\mathbf{X}'\mathbf{y},$$

where $\mathbf{X}'\mathbf{X}\boldsymbol{\rho}_i = \boldsymbol{\lambda}_i$ and with

$$\mathbf{V} = (\boldsymbol{\rho}'_i \boldsymbol{\lambda}_j)$$

it is the case that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{V}\sigma^2)$$

and

$$\frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \mathbf{V}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{ms^2} \sim F_{m, n-r},$$

where $F_{m, n-r}$ follows the F distribution with numerator degrees of freedom equal to m and denominator degrees of freedom equal to $(n - r)$.

(v) Furthermore,

$$\frac{\hat{\boldsymbol{\theta}}' \mathbf{V}^{-1} \hat{\boldsymbol{\theta}}}{ms^2}$$

is distributed according to the noncentral F distribution with the same degrees of freedom and with numerator noncentrality equal to $(\boldsymbol{\theta}' \mathbf{V}^{-1} \boldsymbol{\theta})/\sigma^2$, or half of this with an alternative definition.

(vi) For any sum of squares $SS_i (i = 1, 2, \dots, k)$ in the ANOVA table and $SS(\text{residuals})$ we have

$$\frac{SS_i/r_i}{SS(\text{residuals})/(n - r)} \sim F'_{r_i, n-r},$$

where $F'_{r_i, n-r}$ denotes the non-central F -distribution with r_i and $n - r$ d.f.

The random variables defined in (ii) and (iv) can be used to test hypotheses about a single estimable function and about a set of m estimable functions, respectively. And, finally, the statistic defined in (vi) can be used to test $H_0: \boldsymbol{\nu}'_i \boldsymbol{\nu}_i / \sigma^2 = 0$, that is, the noncentrality parameter for SS_i equals zero ($i = 1, 2, \dots, k$). If H_0 is true, then $F'_{r_i, n-r} = F_{r_i, n-r}$, the central F -distribution.

4.18 MIXED MODELS

4.18.1 The Notion of Fixed, Mixed and Random Models

The general linear models that we have discussed so far are referred to, following terminology introduced by Eisenhart (1947), as *fixed effects models* or, for short, *fixed models*. This means that, rewriting model (4.1) as model (4.74), the individual terms in $\boldsymbol{\beta}^*$ in the model

$$\mathbf{y} = \mu \mathbf{J} + \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{e} \quad (4.74)$$

are (unknown) constants or fixed effects. This is the type of model most often used in explaining observations from experimental studies, as we shall describe in the following chapters.

There are, however, situations where we may want to partition β^* and, conformably, \mathbf{X}^* and rewrite (4.74) as

$$\mathbf{y} = \mu\mathbf{J} + \mathbf{X}_1^*\beta_1^* + \mathbf{X}_2^*\beta_2^* + \mathbf{e} \quad (4.75)$$

where the elements of β_1^* represent fixed effects and the elements of β_2^* are random variables having certain distributional properties. Model (4.75) is then referred to as a *mixed effects model* or, for short, *mixed model*. As an extreme case, model (4.75) with $\beta_2^* = \beta^*$ is referred to as a *random effects model* since all the terms, except μ , in (4.75) are random variables. The other extreme, of course, is $\beta_1^* = \beta^*$, the fixed effects model.

4.18.2 Aitken-like Model

For ease of notation we now rewrite (4.75) as

$$\mathbf{y} = \mu\mathbf{J} + \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{e}, \quad (4.76)$$

where $\mathbf{Z}\gamma$ represents the random part of the model. Since γ and \mathbf{e} represent vectors of random variables with $E(\gamma) = E(\mathbf{e}) = \mathbf{0}$ and

$$\begin{aligned} \text{var}(\gamma) &= E(\gamma\gamma') \equiv \mathbf{V}_\gamma \\ \text{and} \\ \text{var}(\mathbf{e}) &= E(\mathbf{e}\mathbf{e}') \equiv \mathbf{V}_e \end{aligned} \quad (4.77)$$

we can rewrite (4.76) as

$$\mathbf{y} = \mu\mathbf{J} + \mathbf{X}\beta + \mathbf{e}^* \quad (4.78)$$

with $E(\mathbf{e}^*) = \mathbf{0}$ and, using (4.77) and, assuming that γ and \mathbf{e} are uncorrelated,

$$\text{var}(\mathbf{e}^*) = \mathbf{Z}\mathbf{V}_\gamma\mathbf{Z}' + \mathbf{V}_e \equiv \mathbf{V}^*. \quad (4.79)$$

Obviously, \mathbf{V}^* of (4.79) is a real symmetric matrix and assuming that it is invertible it appears that we find ourselves in the situation of Theorem 4.2 (see Section 4.16.2). The difficulty, however, is that \mathbf{V}^* in (4.79) depends, generally, on more than one unknown parameter, namely in particular the variance and covariance components of γ in (4.76).

Under these circumstances we cannot, obviously, obtain the equations (4.67). An easy way out of this dilemma is to estimate the unknown variance and covariance components and substitute these estimates in (4.67). In other words, we estimate \mathbf{V}^* in (4.79) by $\hat{\mathbf{V}}^*$, say, and then solve the Aitken-like equation

$$\mathbf{X}' \left(\hat{\mathbf{V}}^* \right)^{-1} \mathbf{X}\mathbf{b} = \mathbf{X}' \left(\hat{\mathbf{V}}^* \right)^{-1} \mathbf{y}. \quad (4.80)$$

The solution of (4.80) depends, of course, on the type of estimation used to obtain $\hat{\mathbf{V}}^*$, for instance, ANOVA-type estimation. Thus, the solution to (4.80) is no longer BLUE.

4.18.3 Mixed Models in Experimental Design

We conclude this section by giving a brief discussion of the occurrence of mixed models in the context of intervention studies. Generally speaking, they do not occur very often. Referring to (2.3), we recall that the essential parts of the linear model with respect to this question are the treatment effects, the design effects, and possibly treatment \times design interaction effects. The design effects refer mainly to blocking effects as related to intrinsic or nonspecific factors (see Section 2.2.4). To the extent that the levels of a particular blocking factor can be considered to constitute a random sample from a larger population of such levels, the effects of that blocking factor may constitute random effects. As the treatment effects are always fixed effects, the above situation may thus lead to a mixed model, but never to a random effects model. We should emphasize here that if blocking effects are considered to be random effects, then also possible treatment \times blocking interaction effects are also random effects.

In its simplest form a mixed linear model can occur to describe data from a block design. In the model (see (9.5))

$$y_{ik} = \mu + \beta_i + \tau_k + e_{ik}$$

the block effects, β_i , may in certain situations be considered to represent random effects. This is of particular importance for the so-called recovery of inter-block information in incomplete block designs, as described in detail in Section II.1.7. The β_i are considered to be i.i.d. random variables with mean 0 and variance σ_β^2 . Different methods for estimating \mathbf{V}^* in (4.79) are described in Sections II.1.10 and II.1.11.

A somewhat more complicated situation arises when in a block design the block effects are considered to be random effects and block \times treatment interaction is included in the model (see (9.75)), that is, we have

$$y_{ik\ell} = \mu + \beta_i + \tau_k + (\beta\tau)_{ik} + e_{ik\ell}$$

The problem arises with defining the distributional properties of the interaction effects, $(\beta\tau)_{ik}$. A commonly used approach is to consider them to be i.i.d. with mean 0 and variance $\sigma_{\beta\tau}^2$. A different approach based on finite population and randomization theory is presented in Section 9.7.5. Both approaches lead to the same result concerning inferences about the treatment effects, τ_k , but lead to different results concerning inferences about σ_β^2 . For a general discussion of the controversy in the context of observational studies we refer the reader to, for example, Hocking (2003), Lencina et al. (2005), Nelder (1994, 1995) and Voss (1999).

EXERCISES

- 4.1** Prove that the NE $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ is consistent in \mathbf{b} for all \mathbf{y} .
- 4.2** Prove that there exists a $\boldsymbol{\rho}$ such that $\mathbf{X}'\mathbf{X}\boldsymbol{\rho} = \boldsymbol{\lambda}$ implies $\boldsymbol{\lambda}' = \mathbf{a}'\mathbf{X}$ and vice versa.
- 4.3** Prove that the shortest solution of a consistent set of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{b}$.
- 4.4** Verify that \mathbf{A}^- of Example 4.1 satisfies the properties of a generalized inverse.
- 4.5** Verify that \mathbf{A}^+ of Example 4.2 satisfies the properties of a Moore-Penrose inverse.
- 4.6** Prove that the NE for

$$\begin{aligned} \mathbf{y} &\doteq \mathbf{X}\boldsymbol{\beta} \\ \mathbf{C}\boldsymbol{\beta} &= \mathbf{c} \end{aligned}$$

is given by (4.20).

- 4.7** Prove the basic properties (i), (ii), (iii) and (iv) of the NE (4.20).
- 4.8** Refer to Section 4.6.2 and prove that $\mathbf{X}'\mathbf{D}\mathbf{X}$ is s.i.p. and equals $\mathbf{X}[\mathbf{X}(\mathbf{I} - \mathbf{C}^+\mathbf{C})]^+$.
- 4.9** Prove that
- $$\text{rank} \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{C}' \\ \mathbf{C} & \emptyset \end{pmatrix} = \text{rank} \begin{pmatrix} \mathbf{X} \\ \mathbf{C} \end{pmatrix} + \text{rank}(\mathbf{C}).$$
- 4.10** Prove that (4.22), with appropriate conditions on \mathbf{C} , is a solution of (4.20).
- 4.11** Consider the following balanced data structure: We have four factors A , B , C , D , where A and B are crossed, C is nested in AB and D is nested in C .
- (i) Draw a structure diagram.
 - (ii) Give all admissible means.
 - (iii) Write out an identity for the individual observation in terms of components obtained from the admissible means.
 - (iv) Write a model in standard notation.
 - (v) Impose side conditions on the parameters to remove overparameterization.
 - (vi) Write out the ANOVA table.

- 4.12** For the three-way cross-classification write out all possible well-defined models.
- 4.13** (i) Give a definition for connectedness in a three-way cross-classification assuming that all interactions are zero.

- (ii) Consider the three-way classification with factors A, B, C , where each factor has four levels. Denote a design point by (i, j, k) where $i, j, k = 1, 2, 3, 4$ indicate the levels of the factors A, B, C , respectively. Suppose we have the following design points:

$$\begin{array}{cccc} (1, 1, 1), & (1, 2, 2), & (1, 3, 3), & (1, 4, 4) \\ (2, 1, 4), & (2, 2, 1), & (2, 3, 2), & (2, 4, 3) \\ (3, 1, 3), & (3, 2, 4), & (3, 3, 1), & (3, 4, 2) \\ (4, 1, 2), & (4, 2, 3), & (4, 3, 4), & (4, 4, 1) \end{array}$$

Show that the design is connected.

- (iii) Write out and solve the NE for the design in (ii).
- (iv) Write out the model for the design in (ii) in matrix notation and show that the sum of squares for any factor is independent of the order of the factors in the model.
- 4.14** A complete set of linearly independent functions for the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ is a set of $r_{\mathbf{X}}$ linearly independent functions $\lambda'_k \boldsymbol{\beta}$ ($k = 1, 2, \dots, r_{\mathbf{X}}$), where $r_{\mathbf{X}}$ is the rank of \mathbf{X} , such that all estimable functions can be generated from this set.

- (i) For the linear model

$$y_{ij} = \mu + a_i + b_j + e_{ij}$$

($i = 1, 2, \dots, A; j = 1, 2, \dots, B$) obtain a complete set of linearly independent estimable functions.

- (ii) Obtain a complete set of linearly independent estimable functions involving only the a_i ($i = 1, 2, \dots, A$) and show that the sum of squares associated with this set is identical to $\text{SS}(A) = b \sum (\bar{y}_{i.} - \bar{y}_{..})^2$, the usual sum of squares for factor A .

CHAPTER 5

Randomization

5.1 INTRODUCTION

We have seen (see Chapter 4) that if we use the GMNLM

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad \mathbf{e} \sim \text{MVN}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

then we can go through the panorama of conventional statistical ideas, that is, estimation of parametric functions, estimation of error, statistical tests, and statistical intervals.

Insofar as we are merely studying mathematical statistics per se, we have completed the basic ideas. But our interest must surely be directed, in part, at least, to the use of the ideas in the “improvement of natural knowledge.”

What are the problems in applying the mathematical material? First, we have to envisage a population of repetitions. Ordinarily, in substantive experimental science, the population of repetitions over which certain statistical or stochastic properties are to hold is defined by the experimental protocol. This will say something like the following: If you do such and such, then such and such will happen. This is, of course, merely an assertion. In order to make the assertion one will have done “such and such” a number of times, one will have obtained results, and one will demonstrate that the data follow the model or class of models one asserts. Whether this will hold up for a new trial is, of course, mere speculation, hope or faith. The status is no better than our faith that the sun will rise tomorrow morning. This is the classical Humean problem of induction.

5.1.1 Observational versus Intervention Studies

Suppose now that we have observational data. To be specific, suppose we have observed two groups of humans. One group has been taking large doses of vitamin C for five years, and the other group has been taking no vitamin C over and above what is obtained in an “ordinary” diet. Suppose further, that we have a measure of the frequency, duration, and intensity of the common cold for each member of each group.

Our problem is then to set up a model for the data. What population of repetitions are we to assume? What are we to assume about the behavior of deviations from true values in this population of repetitions? The answers to these questions are not at all obvious. We shall surely be involved in model search, and the outcome of our study can have no tighter outcome than the assertions, for which we give, of course, our basis, that the data are “like” a realization of a stochastic model, say $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \sigma^2 \mathbf{I})$, and that given this basis our estimates are such and such, our statistical intervals are such and such, and so on.

The purpose of the remarks is not to denigrate observational studies. There are huge areas of human interest, for instance, astronomy and cosmology, in which we can clearly do no experimentation, though we can experiment on many of the scientific bases for our theories. At a more human and “down to earth” level, we cannot set up an experiment to determine the effects of cigarette smoking in humans. We cannot do an experiment to show that thalidomide produced *in humans* the awful effects we believe it to produce. We do have, in this case, very strong evidence that it does so. And, we can do actual experiments with other organisms, which we have strong reason to believe mimic essentially perfectly what happens in humans.

We now turn from observational studies to experimental (or intervention) studies. We take, for discussion, a “simple” experiment in which the material to be experimented on is humans and the treatments to be compared are interventions to overcome the problems of heart disease. These interventions will be taken to consist of no intervention, cardizem, procordia angioplasty, and bypass surgery. The point of the experiment is that there is a subset of our human population which suffers from heart disease, and the problem is to obtain information on what the effects of the four named interventions will be. We prefer the name intervention experiment rather than the vague general term, experiment. The point of the study is that the experimental units alter over time, and we wish to intervene in the dynamical process of each unit to produce a good outcome.

In attempting to assess the different interventions, we shall call on all the available opinions, especially scientific experience that is available. However, the dynamical situation is so complex that we shall be forced to do an experiment, or, indeed, many experiments.

Obviously, we shall need a number of experimental units. We shall set up a protocol for selecting candidates for experimentation. Obviously, each treatment must be a somewhat appropriate treatment for each candidate.

We suppose that 20 candidates are available. We then have the problem of deciding how to allocate treatments to candidates—with the obvious requirement that each candidate can receive one and only one treatment. Suppose we have decided on a treatment allocation and we have imposed that allocation and conducted our experiment. We suppose that a treatment period has been chosen before the experiment. Then at the end of the experiment we have 20 doublets of, say, (i, t) , where i is the name (or number) of the unit (person) and t is the name (or number) of the treatment, and for each doublet the outcome, $y(i, t)$ say, which we take to be a scalar. Our problem is easy to state: What conclusions can we draw about differences of treatments? It is obvious that treatment effects are confounded with unit effects. The problem is obvious in the case of two doublets (1, 1) and (2, 2) with observations $y(1, 1)$ and $y(2, 2)$. We do not know

what $y(1, 2)$ or $y(2, 1)$ could be. So, if our data triples are $(1, 1, 20)$ and $(2, 2, 10)$, we can conclude only that we cannot determine whether the difference between the yields of 10 and 20 occurs because there is no treatment effect and that unit 1 gives a response greater by 10 than unit 2, or that treatment 2 gives a yield 10 less than treatment 1, or that unit-treatment combination $(1, 1)$ gives a yield of 20 and $(2, 2)$ a yield of 10. How is one to get around this basic indeterminacy?

5.1.2 Historical Controls versus Repetitions

We can, perhaps, call on our past experience and say that a difference of 10 ($= 20 - 10$) has occurred very frequently in observations on the same treatment. Or, we can say, from past experience, that a difference of 10 has never occurred under the same treatment. Perhaps we can condense our historical experience into a probability distribution, an empirical Bayes experience, that the difference between two individuals on the same treatment is distributed with mean μ and variance, 4, say. We would then have to say that we have observed a random variable, O say, with mean μ , the unknown treatment difference, and standard deviation equal to 2. We can then say that $(O - \mu)$ is an approximate pivotal so that, according to Tchebycheff's inequality, $\text{Prob}\{|O - \mu| \geq k2\}$ is less than or equal to $1/k^2$ for any $k \geq 1$.

The procedures of the previous paragraphs are in a general category called "use of historical controls." Undoubtedly, this procedure has been used very widely throughout the development of science, certainly in physics and chemistry. In the absence of appropriate controls, there is no alternative to requiring that a study contain its own controls.

So we now ask if there is an experimental protocol that contains *within itself* a population of repetitions, and such that we may apply our probabilistic models to the resulting data and obtain conclusions, as regards statistical tests and statistical intervals which we may have faith in *because of the experimental protocol we have followed*.

The suggestion of R. A. Fisher (1935) in this respect is the use of randomization. We shall follow to some extent the sequence of ideas that Fisher used.

5.2 THE TEA TASTING LADY

R. A. Fisher wrote his classic, *The Design of Experiments*, in 1935. He opened his exposition with the most famous experiment in statistical thinking, the lady-tasting-tea experiment. The lady of the Rothamsted Experiment Station staff claimed that she could discriminate between a cup of tea made with milk and one with tea added first. Fisher's experiment design consisted of making 8 cups of tea with 4 made in one way and 4 in the other. The lady was told of this structure. The 8 cups are presented to the lady in random order and she has to partition the 8 cups into two sets of 4. The interpretation will be made on the basis that there are $70 [= 8!/(4!4!)]$ partitions. So, if the assignment was random, the probability of the lady obtaining the correct partition is $1/70$, *if she can not discriminate*. Because $1/70$ is a small probability, it is rational to conclude that if she obtains the correct partition she has given evidence in favor of her claim.

Fisher then discusses the experimental technique that must be followed if the probability of $1/70$, under the null hypothesis, can be justified. He mentions temperature of infusion and the nature of the cups, as two possible differences, which would invalidate the probability. He says these are only two possibilities from an indefinitely large number of such. If the 8 cups are prepared and laid out for presentation to the lady in positions 1 to 8, in $8!$ ways, indexed by i and we number the possible partitions by 1 to 70, indexed by j , then there will be probability p_{ij} with $\sum_i p_{ij} = 1$ of the lady choosing partition j . If the correct partition is chosen with probability of $1/70$, the probability of the lady choosing the correct partition is $\frac{1}{70}[\frac{1}{8} \sum_{ij} p_{ij} \frac{1}{70}]$. This probability will hold regardless of the nature of the cups, the method of preparation, and the method of presentation of the cups to the lady. The relevant part of the whole prescription is that the partition used has a probability of $1/70$ regardless of the conduct of the experiment. So, if this partition is obtained after everything else has been done, the probability under the hypothesis of no discrimination ability is $1/70$. This is true even if 4 cups are paper or bone china, or 4 cups inadvertently receive sugar, or whatever.

The tea tasting experiment discussed above has a special structure in that the outcome of the experiment comes about because the taster has to make a comparison of the 8 cups of tea. The taster does not make a quantitative assessment of the properties of each cup. Fisher gives a discussion of the sensitivity of his design which is unsatisfactory, as pointed out by Neyman (1950). Fisher uses a definition: One experiment, E_1 , is more sensitive than another, E_2 , if E_1 “will allow the detection of a lower degree of sensory discrimination, or, in other words, a quantitatively smaller departure from the null hypothesis” (Fisher, 1937, p. 25). For this idea to be implemented, we must be able to compare E_1 and E_2 with respect to the deviations from the null hypothesis that they will express. In an attempt to convince us on Fisher’s ideas, he says that an experiment with 12 cups, 6 of each kind, is more sensitive than the one with 8 cups, 4 of each kind. However, Fisher assumes that a difference from the null hypothesis will be the same in the two experiments. This is obviously not the case, because the task involves comparisons, and making a partition of 12 cups into two groups of 6, is more difficult than making a partition of 8 cups into two groups of 4, when there is the same difference between the two types of tea. In the tea tasting example, the only way to increase sensitivity is to repeat the study with the same design. So, for instance, with two repetitions, the probability of two successes is $1/4900$ and of one success is .028, under the null hypothesis.

5.3 TRIANGULAR EXPERIMENT

It is also useful to mention a much smaller type of discriminatory test, the triangular test. The question is whether a “taster” can discriminate between two versions of a drink. The taster is presented with two specimens of drink A and one of drink B , and is asked to pick out the odd one. In this case, with the proper randomization and experimental technique, the probability of being correct in the absence of detection of a real difference is $1/3$. This test has been used widely in the beer industry, for instance.

5.3.1 Medical Example

Let us now adapt the ideas of the triangular test experiment to a more serious matter. The experiment we shall consider is a very small one, and if considered useful, it should be repeated many times. Suppose we are considering a disease in humans. The disease could be a minor one such as “the common cold” or a very serious one such as cancer. Suppose, furthermore, that we have two possible treatments. In the case of the common cold, these could be

1. Go to bed for three days.
2. Take drug X according to a prescribed regimen for three days.

In the case of cancer the two treatments could be

1. Undergo a regimen of radiation plus a potential anticancer chemical.
2. Do the same as 1 with a different potential anticancer chemical.

The protocol will be that we decide to give one of the two treatments to one of three patients, and the other treatment to the other two patients. We will then have the patients examined by a doctor with a prechosen observation protocol and whatever additional observations he may choose, and this examining doctor is to specify which of the three patients received the odd treatment. The examining doctor will be given, of course, no information on which patients received which treatment. This example is interesting because of several features which contrast with the beer tasting situation. In that case, we can easily envisage obtaining three glasses which are so nearly identical that one glass would not be sufficiently different from the other two to cause it to be chosen as the odd one. In contrast, our three patients will be not “nearly identical” by any stretch of the imagination. They will surely differ in age, in their whole backgrounds, their weights, their dietary intakes, their personalities, and so on. We, the applier of the treatments, may have unconscious biases. The doctor who examines the patients at the end of the experimental period may have unconscious biases. He may think that the appearance of a particular symptom or reaction to treatment is significant. The situation is, formally, analogous to the beer-tasting situation, with the presence of a high amount of data which may be entirely irrelevant, but that may influence the judgment of the evaluator. It is obvious, we suggest, that the randomized triangular test is a candidate design for the study. The end result is, of course very simple. Did the evaluator pick out the odd treatment correctly? To assert the force of the whole experimental design, we adduce the simple probabilistic fact: the probability—and a frequency probability at that—that the evaluator makes the correct choice of the “odd” patient, under the hypothesis that he cannot really discriminate is $1/3$.

5.3.2 Randomization, Probabilities, and Beliefs

Some will attempt to argue, perhaps, that with such a small experiment, randomization is unnecessary. To do so is to fail to see the nature of the logical argument that is being followed. It could be that the evaluator will pick one patient out of the three to be the

odd one for reasons related only to the patients. What then is the probability that he will pick out the odd one? There can be a probability *only* if there is a probability that any particular individual is in fact the odd one vis-à-vis treatment. One cannot determine probabilities without injecting probabilities. One can calculate a probability *only* on the basis of probabilities of elementary sets.

The argumentation above has been criticized particularly by some representatives of the neo-Bayesian school of statistics. Their idea is, it seems, that without the slightest use of randomization in the protocol, the rational person may have the belief probability of $1/3$ that the evaluator will pick out the odd one correctly in the absence of there being a treatment-induced basis for discrimination. An assertion of this sort is absurd. Who can formulate such a belief probability of $1/3$ without an understanding of the psychophysical processes of the evaluator? It would seem that what is being called on is an assumption of ignorance, namely that with three possible decisions the evaluator will make any one of these with a probability of $1/3$. One may mention a variety of problems. The medical evaluator knows, of course, the disease; he may have the idea that fair-haired people are different in their reactions to the disease from dark-haired people, and he may choose the odd one on this basis. In fact, the study of how the evaluator picks out the odd one could be a huge investigation per se. One would have to do the following sort of investigation of the evaluator. One would convince him that we have done a real experiment, when, indeed, we have not. The evaluator would then pick out the odd one, and we could collect a set of data, consisting of attributes of each triple of patients and of the patient he chooses as the odd one. One could then attempt to determine how he does, in fact, pick out the odd one. We close the discussion of this view with the remark that the force of randomization is accepted throughout experimental sciences in which there is unavoidable variability between experimental units.

5.4 SIMPLE ARITHMETICAL EXPERIMENT

5.4.1 Noisy Experiments

We now consider a class of experiments of the following nature. We have N experimental units, which may be mice, men, plots of land, pieces of steel, 8-hour time segments of a functioning chemical reactor, “pieces” of the lower atmosphere, such as clouds, or whatever. We have t treatments, one of which can be imposed on any one experimental unit. A treatment will then be imposed on each experimental unit and after a chosen period an attribute of the experimental unit will be observed, and we suppose this attribute to be an arithmetical or interval measurement, such as height, weight, percent conversion, tensile strength, or whatever. Let us label our units by i ($i = 1, 2, \dots, N$) and our treatments by j ($j = 1, 2, \dots, t$) and our final observation by y_{ij} . Clearly for given i , we shall have only one j represented. We may represent our possible data by a two-way table:

Unit	Treatment			
	1	2	...	t
1				
2				
\vdots				
N				

Now it is obvious that within any row of this table, we shall have only one cell occupied. Suppose that we have an observation y_{13} , that is, unit 1 received treatment 3. Our task is to try to understand the observation y_{13} . Suppose $y_{13} = 21$. Then we know that this is a function of the unit and of the treatment. Consider, for example, the following additive model (see Chapter 6):

$$21 = 5 + 16,$$

by which we mean that the unit value is 5 and the treatment contributed 16. Or, for instance

$$21 = 26 - 5,$$

by which we mean that the unit contributed 26 and the treatment contributed minus 5. Obviously, we cannot distinguish among these possibilities, and indeed we can write

$$21 = u_1 + t_3,$$

and we can see that there is an infinity of values for u_1 , each with an associated value for t_3 which satisfy this equation.

If, of course, one could assert that without treatment, unit 1 would have given an observation of 24, say, one could then assert that the observation $y_{13} = 21$ tells us that t_3 equals minus 3. In non-noisy sciences one may well be able to be confident in taking such a view. But, suppose our observation is weight at age 6 of a child entering the experiment at age 5. Obviously, one cannot be assertive about what a child of age 5 will weigh at age 6. Or to take a harder example, suppose we have a strain of mice or men, such that say, 20% develop a certain disease by a certain age. Then our task, in order to use this sort of argument, is to state what the status of each particular individual will be with respect to the disease at the age of post-experiment observation. In the case of mice or men to consider doing this effectively is simply ludicrous. Contrariwise, if the experimental unit is a carefully prepared test tube with well-defined and well-determined contents, and the question of what the status of that test tube will be in, say, 30 minutes, presents no problems. This latter example is a case from a nonnoisy science. Though we have to realize that the deeper the investigation of *any* scientific field the more noisy it becomes. So the distinction is *not* between different fields of science in their basic natures, but between different fields of science at particular levels of noise. The noise level in basic physics and chemistry which is

now at high school or beginning college level, is very low. In the determination of the weight of an electron it is higher but controllable. In radioactive decay, it may be high and completely uncontrollable.

5.4.2 Investigative Experiments and Beliefs

Let us broaden our consideration of the data we might get. Suppose we have 6 units and 2 treatments and observe

$$y_{12} = 22, \quad y_{21} = 14, \quad y_{32} = 30, \quad y_{41} = 18, \quad y_{51} = 24, \quad y_{62} = 42.$$

We see that the average for treatment 1 is $18\frac{2}{3}$ and for treatment 2 is $31\frac{1}{3}$. Are we then to infer, surmise or guess, that treatment 2 gives a result greater than treatment 1 by $12\frac{2}{3}$? How indeed, may we make such a surmise? We may make the following sort of statement: We have looked at the set of 3 units which received treatment 1 and the set of 3 units which received treatment 2, and *we believe* that these two sets would give nearly the same means if, in fact, there were no treatment effect. Or one might surmise the following: *We believe* that the difference between the two means in the absence of a treatment difference would have been no more than 4; so *we believe* the effect of the treatment is somewhere between $8\frac{2}{3}$ and $16\frac{2}{3}$.

We suggest that while our beliefs should be given some weight, one simply does not know how much weight to give them. We have well-trained and well-intentioned investigators who exhibit a wide panorama of beliefs. The task of investigative experimental science is to remove the role of personal belief that cannot be validated in some way. And the fact that an individual has had a good record of his prior beliefs being sustained by investigation is a weak straw (but in many cases, let it be said, the only straw) on which to base our own outlook. Perhaps some examples, without possible citation, should be given:

- (a) Some years ago, a highly trained and successful medical worker had the very strong belief that stomach ulcers in humans could be cured by freezing the stomach for a period.
- (b) For several years, some high experts had the strong belief that birth control pills were entirely without risk.
- (c) For many years, some scientists with excellent records have had the strong belief that cloud seeding does, in fact, produce rain.
- (d) Some well-trained scientists have the strong belief that the eating of high-cholesterol foods does not increase the risk of heart disease. Other well-trained scientists are convinced of the opposite.
- (e) Some well-trained scientists believe strongly that present-day uses of weed killers and insecticides are not producing a poisonous environment for all organic life. Other very well-trained scientists believe totally the opposite.

To continue with such a list would be rather tedious, but we do suggest to the reader, and particularly any neo-Bayesian who happens to read the above material, to construct his or her own list of problems on which well-trained, seriously-intentioned scientists, who should be given some partial degree of credence, have widely opposing views. It is true, of course, that at the end, for example, after some tens, hundreds, and thousands of years, enough data will have been accumulated to cause scientists to agree. But after that passage of time, other questions will have arisen on which there is the same diversity of belief. So the neo-Bayesian answer that with infinite data, all reasonable people will agree has truth, but we must counter this with the fact that the essence of science is that the questions at issue change over time. The problem is not to reach the correct answer with infinite data; the problem is to make an approach to truth, and to avoid the prejudices and biases that we inevitably accumulate.

5.4.3 Randomized Experiments

How are we to tackle the dilemma of interpretation of the 6 observations we mention above? To some, and indeed to Fisher, the answer is obvious. We must use a randomized design; we are to select from the 6 units at random 3 units which are to receive treatment 1 with the remaining 3 units to receive treatment 2. The basic idea is, after all, rather natural. Underlying our investigation is a 2×6 table of potential (or conceptual) observations:

	Unit						
	1	2	3	4	5	6	Mean
Treatment 1	y_{11}	y_{21}	y_{31}	y_{41}	y_{51}	y_{61}	$\bar{y}_{.1}$
Treatment 2	y_{12}	y_{22}	y_{32}	y_{42}	y_{52}	y_{62}	$\bar{y}_{.2}$

Our task is to form opinions about the difference between the average of the first row and the average of the second row, with the restriction that we can observe only one number in each column. It is natural, then, to select 3 elements from the first row and this then determines a set of 3 elements in the second row. We now do a little elementary mathematics of finite population sampling. Let us consider the total for treatment 1 under the sampling. There are in fact 6 numbers, y_{i1} , $i = 1, 2, \dots, 6$, and we select at random 3 of these. Call the average of the sample \bar{Y}_1 . Then \bar{Y}_1 is a random variable, and we know

$$E(\bar{Y}_1) = \bar{y}_{.1}$$
$$\text{var}(\bar{Y}_1) = \left(\frac{6-3}{6}\right) \sum_i \frac{(y_{i1} - \bar{y}_{.1})^2}{5}.$$

Similarly, with \bar{Y}_2 equal to the average under treatment 2, we have

$$E(\bar{Y}_2) = \bar{y}_{.2}$$

and

$$\text{var}(\bar{Y}_2) = \left(\frac{6-3}{6} \right) \sum_i \frac{(y_{i2} - \bar{y}_{.2})^2}{5}.$$

Here we use the basic formulae for the expectation which is obvious, and variance of a sample mean with sampling from a population of N finite numbers $\{x_i, i = 1, 2, \dots, N\}$ which is

$$\left(\frac{N-n}{N} \right) \sum_i \frac{(x_i - \bar{x})^2}{N-1}.$$

We are not through, however. We cannot use

$$\text{var}(\bar{Y}_1 - \bar{Y}_2) = \text{var}(\bar{Y}_1) + \text{var}(\bar{Y}_2),$$

because the samples under treatment 1 and under treatment 2 are not independent. This gives us a reason to exposit a little more elementary theory of finite sampling. Let $\delta_i = 1$ if unit i receives treatments 1; $= 0$ otherwise ($i = 1, 2, \dots, 6$). Then our estimator of the difference $3(\bar{y}_{.1} - \bar{y}_{.2})$ is equal to

$$\begin{aligned} T &= \delta_1 y_{11} + \delta_2 y_{21} + \dots + \delta_6 y_{61} \\ &\quad - [(1 - \delta_1) y_{12} + (1 - \delta_2) y_{22} + \dots + (1 - \delta_6) y_{62}] \\ &= \sum_i \delta_i (y_{i1} + y_{i2}) - 6\bar{y}_{.2}. \end{aligned} \quad (5.1)$$

Under random sampling we have the following properties:

$$E(\delta_i) = \frac{1}{2}, \quad E(\delta_i^2) = \frac{1}{2}, \quad \text{var}(\delta_i) = \frac{1}{4}; \quad i = 1, 2, \dots, 6 \quad (5.2)$$

$$\begin{aligned} E(\delta_i \delta_{i'}) &= \left(\frac{3}{6} \times \frac{2}{5} \right) = \frac{1}{5} \\ \text{cov}(\delta_i, \delta_{i'}) &= \frac{1}{5} - \frac{1}{4} = -\frac{1}{20} \quad (i \neq i'). \end{aligned} \quad (5.3)$$

We see this because $\delta_i \delta_{i'}$, is 1 or 0 and is 1 only if treatment 1 falls on i , which has probability $\frac{1}{2}$, and then given that treatment 1 falls on i , treatment 1 falls on i' the probability of which is $2/5$. So we have, using (5.1), (5.2), (5.3),

$$\begin{aligned} E(T) &= 3(\bar{y}_{.1} - \bar{y}_{.2}) \\ \text{var}(T) &= \text{var} \left\{ \sum_i \delta_i (y_{i1} + y_{i2}) \right\} \\ &= \frac{1}{4} \sum_i (y_{i1} + y_{i2})^2 - \sum_{\substack{i, i' \\ i \neq i'}} \frac{1}{20} (y_{i1} + y_{i2})(y_{i'1} + y_{i'2}). \end{aligned} \quad (5.4)$$

Let

$$Y_{i.} = y_{i1} + y_{i2} \quad (i = 1, 2, \dots, 6).$$

Then (5.4) can be written as

$$\begin{aligned}
 \text{var}(T) &= \frac{1}{4} \sum_i Y_{i.}^2 - \frac{1}{20} \sum_{\substack{i, i' \\ i \neq i'}} Y_{i.} Y_{i'.} \\
 &= \frac{1}{4} \sum_i Y_{i.}^2 - \frac{1}{20} \left(Y_{..}^2 - \sum_i Y_{i.}^2 \right) \\
 &= \frac{6}{20} \sum_i Y_{i.}^2 - \frac{1}{20} Y_{..}^2 \\
 &= \frac{3}{10} \left\{ \sum_i Y_{i.}^2 - \frac{Y_{..}^2}{6} \right\} = \frac{3}{10} \sum_i \left(Y_{i.} - \frac{Y_{..}}{6} \right)^2. \tag{5.5}
 \end{aligned}$$

Now suppose that treatments have additive effects, that is, the number with unit i and treatment j , which we have denoted by y_{ij} , is made up additively of a unit effect, u_i and a treatment effect τ_j (see also Section 6.3.1); that is,

$$y_{ij} = u_i + \tau_j.$$

Then

$$Y_{i.} = 2u_i + \tau_1 + \tau_2 \tag{5.6}$$

$$Y_{..} = 12u. + 6\tau_1 + 6\tau_2 \tag{5.7}$$

and we see that by substituting (5.6) and (5.7) into (5.5)

$$\begin{aligned}
 \text{var}(T) &= \left(\frac{3}{10} \right) (4) \sum_i (u_i - \bar{u}.)^2 \\
 &= 6 \sum_i (u_i - \bar{u}.)^2 / 5 = 6\sigma^2
 \end{aligned}$$

if we define

$$\sigma^2 = \sum_i (u_i - \bar{u}.)^2 / 5.$$

Finally, since $\bar{Y}_1 - \bar{Y}_2 = \frac{1}{3}T$, it follows that the variance of the difference of the treatment means is

$$\text{var}(\bar{Y}_1 - \bar{Y}_2) = \frac{1}{9}6\sigma^2 = 2(\sigma^2/3).$$

This result is formally the same as we get with the model

$$y_{ij} = \mu + \tau_j + e_{ij}$$

with the e_{ij} having the Gauss-Markov properties.

Also, elementary computation shows that the expectation under randomization of the mean square within treatments is equal to the quantity (or parameter) σ^2 .

A single such small experiment cannot, of course, tell us a lot. We can certainly imagine repeating this experiment a number of times. For each repetition we will have

an estimated mean difference and a standard error, an estimated standard deviation of this estimated difference. Then we will look at the collection of these results. Obviously, we may apply the *Central Limit Theorem* to infer that the average of a number of such experiments will be normally distributed, and we may apply ordinary tests (using normality) on the average, with a standard error based on the variability between the mean differences for the separate experiments.

As regards the assessment of the significance of an observed mean difference in a single such experiment we will use a procedure called the *randomization test procedure*. Before we give a general description of the randomization test we shall trace briefly the ideas of randomization and the resulting test procedure as developed by R. A. Fisher.

5.5 RANDOMIZATION IDEAS FOR INTERVENTION EXPERIMENTS

The test procedure is an outgrowth by Fisher of his discussion entitled “The Arrangement of Field Experiments,” which was published in the *Journal of the Ministry of Agriculture* in 1926. The ideas are expressed in terms of agricultural experiments, naturally, because Fisher was then statistician of Rothamsted Experiment Station. This paper should be read by all students of experimental design. In the field experiment, treatments are applied to plots of land, and the questions considered are how this assignment should be made and how results with different treatments should be evaluated. The reader should realize that the arguments apply to any interventional comparative experiment: for example, on humans, on animals, on engineering material, and so on.

Fisher bases his *whole* argument on the use of tests of significance. He asks (Fisher, 1926, p. 503): “When is a result significant?” and then “What is meant by a valid estimate of error?” He discusses what may be called the use of historical controls, dismissing them in the context of agricultural experiments (erroneously, to some extent, we judge). He states (p. 504): “A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance.” To obtain a valid estimate of error, he advocates the use of replication, that is, the occurrence of the same treatment on different plots. He says that we wish to quantify the differences between plots with different treatments. We do this by obtaining the estimate of error from differences between plots that are treated alike. This estimate “will only be valid if we make sure that in the plot arrangement, pairs of plots treated alike are not distinguishable from pairs of plots treated differently” (p.506). This prescription is, however, not realizable. In the case of the field experiment, in addition to the positions of the plots, each plot will have many attributes, e.g., nature of soil, pH, amount of N, P, and K, and so on. Each plot thus will be representable as a point in a space with a large number of dimensions. The Fisher prescription described above requires “pairs of plots treated alike be not distinguishable from pairs of plots treated differently.” Fisher says: “An experiment either admits a valid estimate of error or it does not: whether it does so or not, depends not on the actual arrangement of plots, but only on the way that arrangement was arrived at” (p. 508). So “If the arrangement

ABBAABBA was arrived at by writing down a succession of sandwiches ABBA, it does not admit of any estimate of certain validity" (p. 508). Furthermore, according to Fisher: "If the same arrangement happened to occur subject to the condition that each pair of strips shall contain an *A* and a *B*, but that which came first shall be decided by the toss of a coin, then a valid estimate may be obtained from the four differences in yield in the four pairs of strips." He continues: "Thus validity of estimation can be guaranteed by appropriate methods of arrangement . . ." (p. 508). Later, he says: "experiments capable of genuine tests of significance can" easily be designed to be very much more accurate than any experiments ordinarily conducted" (p. 508).

We have found the early writings of Fisher discussed above at best obscure and not entirely coherent.

The beginning of Fisher's arguments lies with the use of significance tests. Unfortunately, Fisher never made clear his ideas on this: the obscurity on this has plagued statistics for the past 80 or more years. Fisher was fond of (obscure) classical theory of errors. He did not make a clear distinction between what we call observational studies and interventional (comparative experimental) studies. We are concerned in this book only with the latter.

Fisher made his ideas of 1926 more clear in his book *The Design of Experiments*. In this book, he pursued his ideas on randomization and made the statement: "The purpose of randomization in this, as in the previous experiments exemplified, is to guarantee the validity of the test of significance, this test being based on an estimate of error model possible by replication" (Fisher, 1937, p. 71). Singularly, in connection with the Latin Square design (see Chapter 10) he says: "The purpose of randomization, necessary to ensure the validity of the test of significance applied to the experiment, consists in choosing one at random of the set of squares which can be generated from any chosen arrangement" (Fisher, 1937, p. 80).

Fisher discusses estimation of error and tests of significance by means of analysis of variance. He continues his treatment by use of tests of significance based on the comparison of certain mean squares in the analysis of variance he considers to be appropriate. He continues his discussion with a treatment of "systematic squares" of which the prime example is the Knut Vik square:

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>
<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>

The point of this square and of the name is that the positions occupied by any treatment are given (nearly) by the knight move in chess. Actually, this is not achieved: consider *A* which is in the sequence of cells, (1, 1), (2, 3), (3, 5), (4, 2), (5, 4). The move from (3, 5) to (4, 2) is not a knight move. "In this arrangement, the areas bearing each treatment are nicely distributed over the experimental area. . ." (Fisher, 1937, p. 87). He says: "The total . . . ascribed to treatments and to error" is "independent of the experimental arrangement." He continues his discussion with the remark, "The failure of systematic arrangements came from not recognizing that the function of the experiment was not

only to make an unbiased comparison, but to supply at the same time a valid statement of its significance" (Fisher, 1937, p. 89).

In fact, the Knut Vik square given above is one of two possible, apart from treatment names, so it will be seen that the only test by analysis of variance combined with randomization can give levels of significance of 50% or 100% only.

The way out of the dilemmas was given by Fisher (1937, Section 21) without real understanding. He says: "In these discussions it seems to have escaped recognition that the physical act of randomization, . . . , affords the means . . . of examining the wider hypothesis in which no normality is implied" (Fisher, 1937, p. 51). His procedure is the use of the randomization test, which is totally related to the randomization that was used.

We now turn to this, which in Fisher's words shows the possibility of an independent check on the more expeditious methods in common use.

5.6 GENERAL IDEA OF THE EXPERIMENT RANDOMIZATION TEST

We first describe the particular example of Fisher. He has 15 pairs of cross-fertilized and self-fertilized plants and the differences in yield between the former and the latter. On the assumption that the members of each pair have been applied to pairs of sites at random, the 15 differences would have occurred with equal frequency with a positive or with a negative sign. The observed total difference was 314 and in the total of 2^{15} possible ($=32,768$) arrangements this difference was equaled or exceeded in 863 cases. The difference in absolute magnitude would be exceeded in 1,726 cases, so the significance level by this procedure is $1,726/32,768 = 5.267$ percent. This may be compared with 5 percent given by the normal theory based t test.

Turning now to comparative experiments, we suppose that we have t treatments that are to be compared using N experimental units. We decide that we shall have N_i units for the i th treatment with $\sum_i N_i = N$. We wish to determine the acceptability of the hypothesis that the treatment differences $\{\tau_i - \tau_j = \theta_{ij}\}$. We can, obviously, adjust the observations according to the hypothesis so that all observations arise under treatment 1. For example, if the hypotheses are $\tau_1 - \tau_2 = 3$, $\tau_1 - \tau_3 = 5$, $\tau_1 - \tau_4 = -4$, etc. (with τ_i being the treatment effects) then the original data y_{ui} (u denoting the experimental unit and i the treatments) are adjusted as follows:

$$\begin{aligned} y_{u_1 1}^* &= y_{u_1 1}, & y_{u_2 2}^* &= y_{u_2 2} - 3 \\ y_{u_3 3}^* &= y_{u_3 3} - 5, & y_{u_4 4}^* &= y_{u_4 4} + 4 \end{aligned}$$

etc. We then evaluate the resultant data y_{ui}^* with respect to the null hypothesis that the treatment differences are null.

The logical basis for the test is that if the hypothesis $\{\tau_i - \tau_j = \theta_{ij}\}$ is true, then after the adjustment the data should be a realization of the data we would observe if there were no differences in treatment effects. Hence, we can apply the test of this null hypothesis to the adjusted data.

What test criterion should we use? We suggest that a good criterion is given by least squares. We shall by randomization choose one of a set of experimental plans. We shall compute the test criterion for the actual plan, equal to C_a , then we shall compute the criterion for each of the plans of the set and shall compute the significance level with respect to the hypothesis, $\{\tau_i - \tau_j = \theta_{ij}\}$, as

$$\frac{1}{s}[\text{numbers of } C \geq C_a], \quad (5.8)$$

where s is the logical number of possible plans.

A simple and instructive example of the randomization test for a design discussed above with $t = 2$, $N_1 = N_2 = 4$ is described by Kempthorne (1952, p. 130). Using (5.8) with $s = 70$, the significance level for the randomization was found to be .71 which compares favorably with the significance level of .63 using the usual F -test for the one-way analysis of variance. This is an important point to which we shall return later (see Chapter 6).

5.7 INTRODUCTION TO SUBSEQUENT CHAPTERS

We have given above the basic ideas and philosophy of randomization in a comparative experiment illustrating some basic aspects with the completely randomized design. For further discussion of intervention experiments, randomization, and inference we refer the reader to Kempthorne (1992).

We shall in Chapter 6 give more detailed theory of randomization for the completely randomized design. We shall discuss estimation of differences of treatment effects, estimation of error, and tests of significance. This is in strong agreement with the views of Cox (2006, p. 192) who states:

“Randomization has three roles in applications: as a device for eliminating biases, for example from unobserved explanatory variables and selection effects; as a basis for estimating standard errors; and as a foundation for formally exact significance tests.”

We shall then give some mathematical and simulation results about the approximation of the randomization test by the corresponding F -test in the analysis of variance.

Then we shall repeat the same line of development for increasingly more complicated structures such as the randomized complete block design (Chapter 9), Latin square design (Chapter 10), and designs that are of split-plot nature (Chapter 13). One important feature of this approach will be to show how the physical act of randomization influences the statistical analysis of data from various experimental designs and how, as a consequence, randomization determines what statistical inferences can be drawn from such data and how it should be done.

This Page Intentionally Left Blank

CHAPTER 6

Completely Randomized Design

6.1 INTRODUCTION AND DEFINITION

The simplest error-control design for comparative experiments is the completely randomized design. Its use and usefulness is predicated by the availability of a set of homogeneous experimental units (EU) (for the description of an EU see Section 2.3). The word “homogeneous” in this context should not be interpreted too narrowly. As explained earlier, there do not exist in nature identical EUs and hence homogeneous here means “alike to the extent possible.” Even that phrase is quite relative. The variability among EUs that arise naturally, for instance, humans of a given gender, within a certain age range, with a certain disease, will be much higher than the variability among EUs that have been manufactured, for example, test tubes under controlled conditions. And yet, in both situations the use of a completely randomized design may be quite appropriate. The implications, however, will become evident as we discuss the nature of this design in more detail.

We shall now give the formal definition of the completely randomized design and discuss in subsequent sections the randomization procedure, the derived linear model, tests of hypotheses, and “sample size” considerations.

Suppose we have t treatments and $N = tr$ homogeneous EUs. Let the tr EUs be partitioned randomly with equal probability into t sets of r EUs. Let the t treatments be assigned to the t sets such that the i th treatment is applied to each of the r EUs in the i th set ($i = 1, 2, \dots, t$). This procedure defines the *completely randomized equal replication design* for t treatments. A realization from this protocol is called a *completely randomized equal replication experiment*. In what follows we shall discuss mainly the equal replication situation and hence we shall refer to such a design simply as a completely randomized design (CRD).

It is clear from this definition that one has a randomized design if and only if one has randomized the assignment of the treatments to the EUs. We shall now describe the randomization process more formally and show how such a mathematical characterization leads to the formulation of a linear model and to the analysis of data from such an experiment.

6.2 RANDOMIZATION PROCESS

6.2.1 Use of Random Numbers

The randomization process for the CRD can be described in various ways which are equivalent to the following:

Label the EUs $1, 2, \dots, N$. Make up chips labeled by $k = 1, 2, \dots, N$. Draw a chip after shaking and label it (11). Discard the chip. Draw a chip after shaking and label it (12). Discard the chip. Draw another chip after shaking and label it (13). Continue this process and thereby establish a random correspondence of $1, 2, \dots, N$ with the tr labels $11, 12, \dots, 1r, 21, \dots, 2r, \dots, t1, t2, \dots, tr$. If chip k , and hence EU k , is associated with the label (ij) , apply treatment i to this EU. More precisely, this will be the j th application of treatment i .

This is, of course, equivalent to establishing a random association of the numbers $1, 2, \dots, N = tr$ to the set of tr labels $11, 12, \dots, tr$ by using a table of so-called random numbers. As an example, suppose $N = 24$ and $t = 6, r = 4$. Then with two-digit random numbers we may discard all the numbers except $01, 02, \dots, 24$. This will prove very tedious. Instead, we may discard $00, 97, 98, 99$. We then associate the random numbers $1, 2, 3, 4$ with the EU number 1 , the random numbers $5, 6, 7, 8$ with the EU number 2 , and so on. If we get a repetition of the associated EU number, we ignore it. So if our random numbers are

07, 21, 34, 65, 43, 22, 05, 83, 77, ...

our associated EU numbers are (ignoring repetitions)

2, 6, 9, 17, 11, 6, 2, 21, 20, ...

and the associated labels are then

11, 12, 13, 14, 21, 22, 23, 24, 31, ...

Consequently, EUs 2, 6, 9, 17 receive treatment 1, EUs 11, 21, 20 receive treatment 2, and so on. Even this process will become tedious, and one may have to alter the algorithm at one stage to pick a random member of the unselected subset.

An alternative procedure is to use a computer program such as given, for example, by SAS PROC PLAN (SAS Institute, Inc., 2002–2003), as illustrated in the following example.

EXAMPLE 6.1: For $t = 4, r = 2$ the SAS input statements are given in Table 6.1a. and the actual design (presented in two different forms) is given in Table 6.1b. We note here that the seed number used in Table 6.1a. can be chosen freely. \square

6.2.2 Design Random Variables

The whole randomization process can be expressed mathematically as follows. Let

$$\delta_{ij}^k = \begin{cases} 1 & \text{if EU } k \text{ is associated with label } ij \\ 0 & \text{otherwise,} \end{cases}$$

such that the following probability statements, denoted by $P(\cdot)$, hold:

$$\begin{aligned} P(\delta_{ij}^k = 1) &= 1/N, \\ P(\delta_{ij}^k = 1, \delta_{i'j'}^{k'} = 1) &= 1/N(N-1), \quad k \neq k', \quad (ij) \neq (i'j') \\ P(\delta_{ij}^k = 1, \delta_{i'j'}^{k'} = 1, \delta_{i''j''}^{k''} = 1) &= 1/N(N-1)(N-2), \\ &\quad k, k', k'' \text{ unequal, } (ij), (i'j'), (i''j'') \text{ unequal,} \end{aligned}$$

and so on.

The variables δ_{ij}^k ($k = 1, 2, \dots, N = rt; i = 1, 2, \dots, t; j = 1, 2, \dots, r$) are called *design random variables*. There are $(rt)^2$ such random variables. Obviously, these have many dependencies among them as illustrated in the following example.

EXAMPLE 6.2: Consider $t = 2, r = 2, N = 4$. Then we have

$$\begin{array}{cccc} \delta_{11}^1, & \delta_{12}^1, & \delta_{21}^1, & \delta_{22}^1, \\ \delta_{11}^2, & \delta_{12}^2, & \delta_{21}^2, & \delta_{22}^2, \\ \delta_{11}^3, & \delta_{12}^3, & \delta_{21}^3, & \delta_{22}^3, \\ \delta_{11}^4, & \delta_{12}^4, & \delta_{21}^4, & \delta_{22}^4 \end{array}$$

If, for instance, $\delta_{12}^2 = 1$, that is, EU 2 receives label 12 and hence treatment 1, then $\delta_{11}^2 = \delta_{21}^2 = \delta_{22}^2 = 0$ and also $\delta_{12}^1 = \delta_{12}^3 = \delta_{12}^4 = 0$. This is so because if EU 2 receives label 12, then EU 2 cannot receive any other label and label 12 cannot be associated with any other EU. \square

Hence, we have, in general,

$$\sum_k \delta_{ij}^k = 1, \quad \sum_{ij} \delta_{ij}^k = 1$$

expressing the fact that one EU receives label ij and EU k receives only one such label. This implies, for example,

$$\begin{aligned} P(\delta_{ij}^k = 1, \delta_{i'j'}^{k'} = 1) &= 0 \quad (ij) \neq (i'j') \\ P(\delta_{ij}^k = 1 | \delta_{i'j'}^{k'} = 1) &= 0 \quad (ij) \neq (i'j') \\ P(\delta_{ij}^k = 1 | \delta_{ij}^{k'} = 1) &= 0 \quad k \neq k'. \end{aligned}$$

Probabilities involving three random variables $\delta_{ij}^k, \delta_{i'j'}^{k'}, \delta_{i''j''}^{k''}$ can also be derived easily, and so on. All one needs to recognize is that the $\{\delta_{ij}^k\}$ are simple Bernoulli (0, 1) random variables and that they are identically but not independently distributed. We have a peculiar, but highly structured dependence.

Table 6.1 Randomization Procedure for CRD

```
a.) Input Statements:
proc plan seed=17683;
factors unit=8;
treatments treat=8 cyclic (1 1 2 2 3 3 4 4);
output out=CRD;
title 'COMPLETELY RANDOMIZED DESIGN (t=4, r=2, N=8)';
run;
proc sort out=CRD;
by unit;
run;

proc print;
run;

proc sort out=CRD;
by treat;
run;

proc print;
run;
```

b.) Output:

COMPLETELY RANDOMIZED DESIGN (t=4, r=2, N=8)				
The PLAN Procedure				
Plot Factors				
Factor	Select	Levels	Order	
unit	8	8	Random	
Treatment Factors				
Factor	Select	Levels	Order	Initial Block / Increment
treat	8	8	Cyclic	(1 1 2 2 3 3 4 4) / 1
-----unit-----				
-----treat-----				
6 2 3 8 7 4 1 5				
1 1 2 2 3 3 4 4				

Table 6.1 (Continued)

Obs	unit	treat
1	1	4 [h]
2	2	1
3	3	2
4	4	3
5	5	4
6	6	1
7	7	3
8	8	2

Obs	unit	treat
1	2	1
2	6	1
3	3	2
4	8	2
5	4	3
6	7	3
7	1	4
8	5	4

We shall now use the mathematical formulation and the statistical properties of the randomization procedure to derive a linear model for the observations from a CRD together with their distributional properties, following Kempthorne (1952; 1955).

6.3 DERIVED LINEAR MODEL

6.3.1 Conceptual Responses and Observations

If EU k receives the label ij , then treatment i is applied to EU k . At the end of the experimental period an observation is made which we denote by y_{ij} . It is, in fact, the observation on the j th occurrence of the i th treatment. We may suppose that if treatment i is applied to EU k the true (or conceptual) response is a number, T_{ik} say. We suppose that if we could, in fact, impose every treatment on every EU, we could observe the totality of numbers $\{T_{ik}\}$. But we cannot do this as we can apply only one treatment to each EU and that is determined by the randomization process. Using the

design random variables, δ_{ij}^k , we can then link the y_{ij} and the T_{ik} as

$$y_{ij} = \sum_{k=1}^N \delta_{ij}^k T_{ik}. \quad (6.1)$$

This states that if EU k receives the label ij , then we observe T_{ik} . The $\{T_{ik}\}$ are fixed numbers under repetitions of the randomization. This is just a $t \times N$ array of numbers, and we may write the identity

$$T_{ik} = \bar{T}.. + (\bar{T}_{i.} - \bar{T}..) + (\bar{T}_{.k} - \bar{T}..) + (T_{ik} - \bar{T}_{i.} - \bar{T}_{.k} + \bar{T}..), \quad (6.2)$$

where $\bar{T}..$ is the overall average of the T_{ik} ; $\bar{T}_{i.}$ is the average of all conceptual responses for the i th treatment; $\bar{T}_{.k}$ is the average of all conceptual responses for EU k .

We shall now assume that

$$T_{ik} = T_i + U_k, \quad (6.3)$$

that is, the response of treatment i applied to EU k is made up additively from a contribution due to the i th treatment, T_i , and a contribution due to the k th EU, U_k . We refer to this as *additivity in the strict sense*. It follows then from (6.3) that

$$T_{ik} - \bar{T}_{i.} - \bar{T}_{.k} + \bar{T}.. = 0$$

and hence (6.2) reduces to

$$T_{ik} = \bar{T}.. + (\bar{T}_{i.} - \bar{T}..) + (\bar{T}_{.k} - \bar{T}..)$$

which, using (6.3), we rewrite as

$$T_{ik} = (\bar{T}.. + \bar{U}..) + (T_i - \bar{T}..) + (U_k - \bar{U}..) \quad (6.4)$$

with $\bar{T}..$ being the average of all treatment contributions T_i , and $\bar{U}..$ being the average of all EU contributions U_k . Letting

$$\mu = \bar{T}.. + \bar{U}.., \quad \tau_i = T_i - \bar{T}.., \quad u_k = U_k - \bar{U}..$$

we rewrite (6.4) as

$$T_{ik} = \mu + \tau_i + u_k. \quad (6.5)$$

It follows immediately from the definitions that

$$\sum_{i=1}^t \tau_i = 0, \quad \sum_{k=1}^N u_k = 0.$$

Substituting (6.5) into (6.1) we obtain

$$y_{ij} = \sum_k \delta_{ij}^k (\mu + \tau_i + u_k) = \mu + \tau_i + \sum_k \delta_{ij}^k u_k, \quad (6.6)$$

where we have used the property that $\sum_k \delta_{ij}^k = 1$. Let

$$\omega_{ij} = \sum_k \delta_{ij}^k u_k. \quad (6.7)$$

Then, finally, y_{ij} can be written as

$$y_{ij} = \mu + \tau_i + \omega_{ij}. \quad (6.8)$$

We refer to (6.8) as the *derived linear model* associated with the CRD.

6.3.2 Distributional Properties

In (6.8) the only random variable on the right-hand side is ω_{ij} . Its distributional properties are determined entirely by those of the δ_{ij}^k . Denoting the expectation operator E and the variance and covariance operator var and cov under the randomization model as E_R , var_R , and cov_R , respectively, we obtain first

$$\begin{aligned} E_R(\delta_{ij}^k) &= \frac{1}{N} \\ \text{var}_R(\delta_{ij}^k) &= E_R((\delta_{ij}^k)^2) - (E_R(\delta_{ij}^k))^2 \\ &= \frac{1}{N} - \left(\frac{1}{N}\right)^2 \\ &= \frac{1}{N} \left(1 - \frac{1}{N}\right) \\ \text{cov}_R(\delta_{ij}^k, \delta_{i'j'}^{k'}) &= -\frac{1}{N^2} & k = k', (ij) \neq (i'j') \\ &= -\frac{1}{N^2} & k \neq k', (ij) = (i'j') \\ &= \frac{1}{N^2(N-1)} & k \neq k', (ij) \neq (i'j'). \end{aligned}$$

Using these results we obtain further

$$\begin{aligned} E_R(\omega_{ij}) &= \sum_k E_R(\delta_{ij}^k) u_k \\ &= \frac{1}{N} \sum_k u_k = 0, \\ \text{var}_R(\omega_{ij}) &= \sum_k \text{var}_R(\delta_{ij}^k) u_k^2 + \sum_{k \neq k'} \text{cov}_R(\delta_{ij}^k, \delta_{ij}^{k'}) u_k u_{k'} \\ &= \frac{1}{N} \left(1 - \frac{1}{N}\right) \sum_k u_k^2 - \frac{1}{N^2} \sum_{k \neq k'} u_k u_{k'} \\ &= \left[\frac{1}{N} \left(1 - \frac{1}{N}\right) \frac{1}{N^2} \right] \sum_k u_k^2 \\ &= \frac{1}{N} \sum_k u_k^2 \end{aligned} \quad (6.9)$$

since $\sum_{k \neq k'} u_k u_{k'} = -\sum_k u_k^2$. Defining

$$\sigma_u^2 = \sum_k u_k^2 / (N - 1) \quad (6.10)$$

we then write (6.9) as

$$\text{var}_R(\omega_{ij}) = \left(1 - \frac{1}{N}\right) \sigma_u^2. \quad (6.11)$$

Similarly, we find for $ij \neq i'j'$

$$\text{cov}_R(\omega_{ij}, \omega_{i'j'}) = -\frac{1}{N} \sigma_u^2. \quad (6.12)$$

Recall that $u_k = U_k - \bar{U}$, denotes the deviation of the contribution of the k th EU from the average contribution of all EUs. Then σ_u^2 can be interpreted as a measure of the variability among the EUs, that is, the heterogeneity of the EUs. Also, ω_{ij} may then be referred to as *unit error*.

We note here parenthetically that the above covariance structure is interesting as an example for which the simple least squares estimators are best unbiased estimators (see Section 4.16.3).

From the results above and the model (6.8) we can easily derive

$$\begin{aligned} E_R(y_{ij}) &= \mu + \tau_i \\ E_R(\bar{y}_{i.}) &= \mu + \tau_i \end{aligned}$$

with $\bar{y}_{i.} = \frac{1}{r} \sum_j y_{ij}$,

$$\begin{aligned} \text{var}_R(y_{ij}) &= \text{var}_R(\omega_{ij}) = \left(1 - \frac{1}{N}\right) \sigma_u^2 \\ \text{var}_R(\bar{y}_{i.}) &= \frac{1}{r^2} \left[r \left(1 - \frac{1}{N}\right) \sigma_u^2 - r(r-1) \frac{1}{N} \sigma_u^2 \right] \\ &= \frac{1}{r} \left(1 - \frac{r}{N}\right) \sigma_u^2 \end{aligned} \quad (6.13)$$

$$\text{cov}_R(\bar{y}_{i.}, \bar{y}_{i'.}) = -\frac{1}{N} \sigma_u^2 \quad (i \neq i'). \quad (6.14)$$

If we consider a contrast among the treatment effects τ_i , say $\sum_i c_i \tau_i$ with $\sum_i c_i = 0$, we find immediately that $\sum_i c_i \bar{y}_{i.}$ is an unbiased estimator for that contrast, that is,

$$E_R \left(\sum_i c_i \bar{y}_{i.} \right) = \sum_i c_i \tau_i$$

with

$$\text{var}_R \left(\sum_i c_i \bar{y}_{i.} \right) = \sum_i c_i^2 \frac{\sigma_u^2}{r} \quad (6.15)$$

using (6.13) and (6.14).

6.3.3 Additivity in the Broad Sense

Our discussion up to this point has been based entirely on the model (6.8) under the assumption of additivity in the strict sense. We have mentioned earlier (see Chapter 2) that associated with each observation are two error components: experimental error and observational error. The only error we have encountered so far is the unit error ω_{ij} associated with the observation y_{ij} . The unit error is part of the experimental error, but in order to incorporate other error components we must broaden our model and our assumptions. Let

$$Y_{ik} = T_{ik} + M_{ik} = T_i + U_k + M_{ik} \quad (6.16)$$

denote the conceptual observation from EU k to which treatment i has been applied. We shall refer to (6.16) as the model under *additivity in the broad sense*. The component M_{ik} expresses what we might call technical error (Wilk and Kempthorne, 1956). This includes:

- (i) treatment error; that is, error due to our inability to replicate a treatment from one application to the next;
- (ii) state error; that is, error due to random changes in the physical state of an EU;
- (iii) selection error; that is, error due to the random selection of EUs for the experiment;
- (iv) measurement error; that is, error due to imprecision in our measurement or scoring procedure;
- (v) sampling error; that is, error due to the random selection of observational units (OUs) for the investigation.

We may consider the errors (i), (ii) and (iii) as part of experimental error, and errors (iv) and (v) as observational error. Accordingly, it is convenient to partition M_{ik} as

$$M_{ik} = E_{ik} + O_{ik} \quad (6.17)$$

to reflect these two components. Using (6.17) and (6.5) we then rewrite (6.16) as

$$Y_{ik} = \mu + \tau_i + u_k + E_{ik} + O_{ik}. \quad (6.18)$$

This is now the conceptual response of applying treatment i to EU k . The actual observation y_{ij} can then be modelled, following (6.1), as

$$\begin{aligned} y_{ij} &= \sum_k \delta_{ij}^k Y_{ik} \\ &= \mu + \tau_i + \sum_k \delta_{ij}^k u_k + \sum_k \delta_{ij}^k E_{ik} + \sum_k \delta_{ij}^k O_{ik} \\ &= \mu + \tau_i + \omega_{ij} + \nu_{ij} + \eta_{ij}, \end{aligned} \quad (6.19)$$

say, where

$$\begin{aligned}\nu_{ij} &= \sum_k \delta_{ij}^k E_{ik}, \\ \eta_{ij} &= \sum_k \delta_{ij}^k O_{ik}.\end{aligned}$$

As an illustration of the error structure as described above we consider the following example.

EXAMPLE 6.3: Suppose we want to compare different spraying regimens for peach trees in an effort to increase the yield and improve the quality of the fruit. We have available an orchard consisting of trees of the same variety and same age. For each of the t regimens we randomly select r trees. The trees are then sprayed at a specified rate on several specified occasions throughout the growing season. We can then identify the various error components as follows:

- (i) treatment error: even though the rate is specified for each tree and occasion the rate may not be achieved exactly and/or the spray may not cover the tree uniformly;
- (ii) state error: the trees may grow differently during the growing season due to different micro climates such as wind, moisture, sun exposure;
- (iii) selection error: different trees could have been included in the experiment;
- (iv) measurement error: the judgement in assessing the quality of the individual peaches may not be quite uniform;
- (v) sampling error: typically only a few peaches per tree are judged for quality, they are picked at random and hence different peaches could have been selected. \square

6.3.4 Error Structure

The quantities E_{ik} and O_{ik} are random variables with mean zero. Furthermore, the δ_{ij}^k , E_{ik} , and O_{ik} are statistically independent. Hence

$$E(\nu_{ij}) = E(\eta_{ij}) = 0$$

and

$$\begin{aligned}\text{var}(\nu_{ij}) &= \sigma_\nu^2 \\ \text{var}(\eta_{ij}) &= \sigma_\eta^2,\end{aligned}$$

say. It then follows that

$$E(y_{ij}) = \mu + \tau_i$$

and

$$\text{var}(y_{ij}) = \left(1 - \frac{1}{N}\right) \sigma_u^2 + \sigma_v^2 + \sigma_\eta^2.$$

Furthermore, as an extension of (6.13) and (6.14) we have

$$\begin{aligned} \text{var}(\bar{y}_{i.}) &= \frac{1}{r} \left[\left(1 - \frac{r}{N}\right) \sigma_u^2 + \sigma_v^2 + \sigma_\eta^2 \right] \\ \text{cov}(\bar{y}_{i.}, \bar{y}_{i' .}) &= -\frac{1}{N} \sigma_u^2 \quad (i \neq i') \end{aligned}$$

so that, corresponding to (6.15), we find, for $\sum_i c_i = 0$,

$$\text{var} \left(\sum_i c_i \bar{y}_{i.} \right) = \sum_i c_i^2 (\sigma_u^2 + \sigma_v^2 + \sigma_\eta^2) / r. \quad (6.20)$$

Expressions (6.15) and (6.20) for $\text{var}(\sum c_i \bar{y}_{i.})$ are the same as would have been obtained if the ω_{ij} were uncorrelated with variance σ_u^2 . We shall, therefore, from now on, mainly to facilitate the computation of variances and other functions of the observations y_{ij} , treat the ω_{ij} as if they were independently, identically distributed (i.i.d.) random variables with mean zero and variance σ_u^2 . Since ω_{ij} and ν_{ij} are components of experimental error it is then also convenient to combine them into one term and define the random variable

$$\varepsilon_{ij} = \omega_{ij} + \nu_{ij}$$

to be the *experimental error* with

$$E(\varepsilon_{ij}) = 0$$

and

$$\text{var}(\varepsilon_{ij}) = \sigma_\varepsilon^2 = \sigma_u^2 + \sigma_v^2,$$

that is, we may consider the ε_{ij} , for purely practical reasons, also as i.i.d. random variables. It follows then from (6.19) that

$$E(y_{ij}) = \mu + \tau_i$$

and from our earlier discussion that

$$\text{var}(y_{ij}) = \sigma_\varepsilon^2 + \sigma_\eta^2,$$

where σ_ε^2 is referred to as the *experimental error variance component*, and σ_η^2 as the *observational (sampling) error variance component*. To condense the notation even further we shall find it usually convenient to use a single error term

$$e_{ij} = \varepsilon_{ij} + \eta_{ij}$$

with

$$\text{var}(e_{ij}) = \sigma_e^2 = \sigma_\varepsilon^2 + \sigma_\eta^2. \quad (6.21)$$

6.3.5 Summary of Results

We can summarize our discussion up to this point as follows:

- (i) Under the assumption of additivity in the strict sense, the randomization process leads to the derived linear model

$$\begin{aligned}
 y_{ij} &= \sum_k \delta_{ij}^k T_{ik} \\
 &= \sum_k \delta_{ij}^k (T_i + U_k) \\
 &= \mu + \tau_i + \sum_k \delta_{ij}^k u_k \\
 &= \mu + \tau_i + \omega_{ij}.
 \end{aligned}$$

- (ii) The ω_{ij} have mean zero and a simple covariance structure

$$\text{cov}_R(\omega_{ij}, \omega_{i'j'}) = \begin{cases} (1 - \frac{1}{N}) \sigma_u^2 & (ij) = (i'j') \\ -\frac{1}{N} \sigma_u^2 & (ij) \neq (i'j') \end{cases}$$

- (iii) A treatment contrast $\sum c_i \tau_i$ is estimated unbiasedly by the same contrast in the treatment means, that is, $\sum c_i \bar{y}_{i.}$ with

$$\text{var}_R \left(\sum_i c_i \bar{y}_{i.} \right) = \sum_i c_i^2 \frac{\sigma_u^2}{r}.$$

- (iv) Under the assumption of additivity in the broad sense, the model in (i) is amended by technical error components to a partly derived, partly assumed model

$$\begin{aligned}
 y_{ij} &= \sum_k \delta_{ij}^k Y_{ik} \\
 &= \sum_k \delta_{ij}^k (T_i + U_k + E_{ik} + O_{ik}) \\
 &= \mu + \tau_i + \omega_{ij} + \nu_{ij} + \eta_{ij}
 \end{aligned}$$

with

$$\begin{aligned}
 E(y_{ij}) &= \mu + \tau_i \\
 \text{cov}(y_{ij}, y_{i'j'}) &= \begin{cases} (1 - \frac{1}{N}) \sigma_u^2 + \sigma_\nu^2 + \sigma_\eta^2 & (ij) = (i'j') \\ -\frac{1}{N} \sigma_u^2 & (ij) \neq (i'j') \end{cases}.
 \end{aligned}$$

- (v) As an extension of (iii), for $\sum_i c_i = 0$,

$$\begin{aligned}
 E \left(\sum c_i \bar{y}_{i.} \right) &= \sum c_i \tau_i \\
 \text{var} \left(\sum c_i \bar{y}_{i.} \right) &= \sum c_i^2 (\sigma_u^2 + \sigma_\nu^2 + \sigma_\eta^2) / r
 \end{aligned}$$

- (vi) Because of the result in (iii) and (v) we treat, in the appropriate context, ω_{ij} as if they were i.i.d. $(0, \sigma_u^2)$ and write the experimental error as

$$\varepsilon_{ij} = \omega_{ij} + \nu_{ij}$$

with

$$\begin{aligned} E(\varepsilon_{ij}) &= 0 \\ \text{var}(\varepsilon_{ij}) &= \sigma_u^2 + \sigma_\nu^2 = \sigma_\varepsilon^2 \\ \text{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) &= 0 \quad (ij) \neq (i'j'). \end{aligned}$$

- (vii) The overall error, experimental and observational, is

$$e_{ij} = \varepsilon_{ij} + \eta_{ij}$$

with

$$\begin{aligned} E(e_{ij}) &= 0 \\ \text{var}(e_{ij}) &= \sigma_\varepsilon^2 + \sigma_\eta^2 = \sigma_e^2 \end{aligned}$$

and, as explained above, we can treat the e_{ij} such that

$$\text{cov}(e_{ij}, e_{i'j'}) = 0 \quad (ij) \neq (i'j').$$

- (viii) Useful expressions for $\text{var}(\sum c_i \bar{y}_{i.})$ are

$$\begin{aligned} \text{var} \left(\sum_i c_i \bar{y}_{i.} \right) &= \sum_i c_i^2 (\sigma_\varepsilon^2 + \sigma_\eta^2) / r \\ &= \sum_i c_i^2 \sigma_e^2 / r. \end{aligned}$$

To make further inferences about treatment comparisons beyond point estimation we need to consider questions of interval estimation or tests of significance (see Sections 6.5–6.7).

6.4 ANALYSIS OF VARIANCE

6.4.1 Deriving the ANOVA Table

As indicated earlier, one of the most important tools for analyzing data from designed experiments is the analysis of variance. For data from a CRD, the analysis of variance (ANOVA) is based on the model

$$y_{ij} = \mu + \tau_i + e_{ij} \tag{6.22}$$

as developed in Sections 6.3.3 and 6.3.4. In linear models terminology this is a one-way classification model and the analysis of variance table is as given in Table 6.2 (see also Section 4.12).

We shall comment in some detail, mainly to lay the groundwork for future chapters, on the ANOVA Table 6.2 and its various parts.

The number of entries (sources) is determined by the number of components, apart from μ , in the model. In this case the model is given by (6.22) and contains two terms, due to treatments (τ_i) and error (e_{ij}). We note here that even if we had written (6.22) in its more explicit form (6.19), the number of entries in the ANOVA table would have been two, since for an actual data set the various error terms cannot be separated (see also Section 2.6). The corresponding partition of the (corrected) total sum of squares, $SS(\text{Total})$ can be obtained by writing the following identity:

$$y_{ij} = y_{..} + (\bar{y}_i - \bar{y}_{..}) + (y_{ij} - \bar{y}_i.)$$

or

$$(y_{ij} - \bar{y}_{..}) = (\bar{y}_i - \bar{y}_{..}) + (y_{ij} - \bar{y}_i.). \quad (6.23)$$

Squaring both sides of (6.23) and summing over both subscripts, we obtain, using the fact that $\sum_{ij}(y_i. - \bar{y}_{..})(y_{ij} - \bar{y}_i.) = 0$,

$$\sum_{ij}(y_{ij} - \bar{y}_{..})^2 = \sum_{ij}(\bar{y}_i - \bar{y}_{..})^2 + \sum_{ij}(y_{ij} - \bar{y}_i.)^2$$

or

$$\sum_{ij}(y_{ij} - \bar{y}_{..})^2 = r \sum_i(\bar{y}_i - \bar{y}_{..})^2 + \sum_{ij}(y_{ij} - \bar{y}_i.)^2, \quad (6.24)$$

which is indeed

$$SS(\text{Total}) = SS(T) + SS(E),$$

where $SS(T)$ and $SS(E)$ refer to the treatment and error sum of squares, respectively.

The partition (6.24) exhibits two things:

- (i) If we substitute in each term on the right-hand side of (6.24) for y_{ij} the model (6.22), we recognize that $\sum_{ij}(y_{ij} - \bar{y}_i.)^2$ is a quadratic function in the e_{ij} only and $\sum_i(\bar{y}_i - \bar{y}_{..})^2$ is a quadratic function in the τ_i and the e_{ij} , hence the names $SS(\text{Error})$ and $SS(\text{Treatments})$, respectively.
- (ii) Since $\sum_i(\bar{y}_i - \bar{y}_{..}) = 0$, this sum contains only $t - 1$ (mathematically) independent terms which accounts for the $t - 1$ degrees of freedom (d.f.) associated with $SS(T)$. Similarly, in $\sum_i[\sum_j(y_{ij} - \bar{y}_i.)]$ we have for every i , $\sum_j(\bar{y}_i - \bar{y}_i.) = 0$. Hence each such sum contains $r - 1$ independent terms and hence the number of d.f. associated with $SS(E)$ is $t(r - 1)$.

Table 6.2 ANOVA for CRD

Source	d.f.	SS	MS	E(MS)	
				add. strict sense	add. broad sense
Treatments	$t - 1$	$r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = SS(T)$	$MS(T) = SS(T)/(t - 1)$	$\sigma_u^2 + \frac{r}{t-1} \sum \tau_i^2$	$\sigma_e^2 + \frac{r}{t-1} \sum \tau_i^2$
Error	$t(r - 1)$	$\sum_{i,j} (y_{ij} - \bar{y}_{i.})^2 = SS(E)$	$MS(E) = SS(E)/t(r - 1)$	σ_u^2	σ_e^2
Total	$tr - 1$	$\sum_{i,j} (y_{ij} - \bar{y}_{..})^2$			

Turning to the expected mean squares, $E(\text{MS})$, the reader will notice that we have given two forms: one based on the assumption of additivity in the strict sense [model (6.8)], and the other based on the assumption of additivity in the broad sense [models (6.19) and (6.22)]. From a practical point of view, that is, for real data sets from a CRD, only the latter is important, but by exhibiting both forms we want to demonstrate their similarity and show that whether we use the covariance structure (6.11) and (6.12) of the ω_{ij} or treat the ω_{ij} together with the ν_{ij} and η_{ij} as i.i.d. random variables, we obtain equivalent results, that is, σ_u^2 in the first form is simply substituted by $\sigma_e^2 = \sigma_u^2 + \sigma_v^2 + \sigma_\eta^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$. This may be a subtle and philosophical point, but it is an important one in the transition from purely derived models to partly derived, partly assumed models.

6.4.2 Obtaining Expected Mean Squares

There are different methods of obtaining $E(\text{MS})$. One is to substitute for the y s in the expression for the mean square the linear model and then evaluate the expected value of that expression. We shall illustrate this for $E[\text{MS}(T)]$ under model (6.8). We have

$$\text{SS}(T) = r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = r \sum_i (\tau_i + \bar{\omega}_{i.} - \bar{\tau}_{.} - \bar{\omega}_{..})^2. \quad (6.25)$$

Now,

$$\bar{\tau}_{.} = \frac{1}{t} \sum_i \tau_i = 0$$

and

$$\begin{aligned} \bar{\omega}_{..} &= \frac{1}{tr} \sum_{ij} \omega_{ij} = \frac{1}{tr} \sum_{ij} \sum_k \delta_{ij}^k u_k \\ &= \frac{1}{tr} \sum_k \left(\sum_{ij} \delta_{ij}^k \right) u_k \\ &= \frac{1}{tr} \sum_k u_k = 0. \end{aligned}$$

Then, (6.25) apart from the multiplier r can be expanded as follows:

$$\begin{aligned}
\sum_i (\tau_i + \bar{\omega}_{i.})^2 &= \sum_i \left(\tau_i + \frac{1}{r} \sum_j \omega_{ij} \right)^2 \\
&= \sum_i \left(\tau_i + \frac{1}{r} \sum_j \sum_k \delta_{ij}^k u_k \right)^2 \\
&= \sum_i \tau_i^2 + \frac{1}{r^2} \sum_i \left(\sum_j \sum_k \delta_{ij}^k u_k \right)^2 + \frac{2}{r} \sum_{ij} \sum_k \tau_i \delta_{ij}^k u_k \\
&= \sum_i \tau_i^2 + \frac{1}{r^2} \sum_{ij} \sum_k \delta_{ij}^k u_k^2 + \frac{1}{r^2} \sum_i \sum_{j \neq j'} \sum_k \delta_{ij}^k \delta_{ij'}^k u_k^2 \\
&\quad + \frac{1}{r^2} \sum_i \sum_j \sum_{k \neq k'} \delta_{ij}^k \delta_{ij}^{k'} u_k u_{k'} + \frac{1}{r^2} \sum_i \sum_{j \neq j'} \sum_{k \neq k'} \delta_{ij}^k \delta_{ij'}^{k'} u_k u_{k'} \\
&\quad + \frac{2}{r} \sum_{ij} \sum_k \tau_i \delta_{ij}^k u_k. \tag{6.26}
\end{aligned}$$

Taking the expected value of the right-hand side of (6.26) and using the properties of the δ_{ij}^k , we obtain

$$\begin{aligned}
E_R \left\{ \sum_i (\tau_i + \bar{\omega}_{i.})^2 \right\} &= \sum_i \tau_i^2 + \frac{1}{r^2} tr \cdot \frac{1}{N} \sum_k u_k^2 \\
&\quad + \frac{1}{r^2} tr(r-1) \frac{1}{N(N-1)} \sum_{k \neq k'} u_k u_{k'} \\
&= \sum_i \tau_i^2 + \frac{1}{r^2} \sum_k u_k^2 - \frac{1}{r^2} (r-1) \frac{1}{N-1} \sum_k u_k^2 \\
&= \sum_i \tau_i^2 + \frac{1}{r^2} [N-1-r+1] \frac{1}{N-1} \sum_k u_k^2 \\
&= \sum_i \tau_i^2 + \frac{1}{r} (t-1) \sigma_u^2,
\end{aligned}$$

where we have used the fact that $N = tr$, $\sum u_k = 0$, and $\sum u_k^2 / (N-1) = \sigma_u^2$. It then follows that

$$E_R\{\text{MS}(T)\} = E \left\{ r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 / (t-1) \right\} = \frac{r}{t-1} \sum \tau_i^2 + \sigma_u^2$$

as given in Table 6.2.

Another method to find $E(\text{MS})$ is to use the fact that for any random variable X ,

$$E(X^2) = \text{var}(X) + [E(X)]^2.$$

We shall illustrate this method also for $E[\text{MS}(T)]$ but this time under model (6.19). To this end, we consider

$$E\{(\bar{y}_{i.} - \bar{y}_{..})^2\} = \text{var}(\bar{y}_{i.} - \bar{y}_{..}) + [E(\bar{y}_{i.} - \bar{y}_{..})]^2.$$

Now

$$\begin{aligned} \text{var}(\bar{y}_{i.} - \bar{y}_{..}) &= \text{var} \left\{ \left(1 - \frac{1}{t}\right) \bar{y}_{i.} - \frac{1}{t} \sum_{i' \neq i} \bar{y}_{i'.} \right\} \\ &= \left(1 - \frac{1}{t}\right)^2 \frac{\sigma_e^2}{r} + \frac{1}{t^2} (t-1) \frac{\sigma_e^2}{r} \\ &= \frac{t-1}{t} \frac{\sigma_e^2}{r} \end{aligned} \quad (6.27)$$

and

$$E(\bar{y}_{i.} - \bar{y}_{..}) = \tau_i, \quad (6.28)$$

where we have used the fact that the e_{ij} can be treated as i.i.d. with mean zero and variance σ_e^2 . It follows then immediately from (6.27) and (6.28) that

$$E\{\text{MS}(T)\} = \sigma_e^2 + \frac{r}{t-1} \sum_i \tau_i^2,$$

as given in Table 6.2.

The expected value of $\text{MS}(E)$ can be obtained similarly. It follows then from Table 6.2 that an estimator for σ_e^2 is

$$\hat{\sigma}_e^2 = \text{MS}(E). \quad (6.29)$$

And hence the standard error for the estimator of the treatment contrast $\sum c_i \tau_i$ is given by

$$\begin{aligned} \text{s.e.} \left(\sum c_i \hat{\tau}_i \right) &= \text{s.e.} \left(\sum c_i \bar{y}_{i.} \right) \\ &= \left(\sum_i c_i^2 \frac{\hat{\sigma}_e^2}{r} \right)^{1/2} \\ &= \left(\sum_i c_i^2 \frac{\text{MS}(E)}{r} \right)^{1/2}. \end{aligned} \quad (6.30)$$

As has been shown in Chapter 4, the ANOVA table can be used to test hypotheses about parametric functions in the context of the underlying linear model (see also Chapter 7). We shall examine the ideas of testing hypotheses about treatment effects using observations from the CRD in the next section.

6.5 STATISTICAL TESTS

The mathematical description of the randomization process together with the assumption of treatment-unit additivity in the strict sense, has allowed us to derive a linear model the properties of which are determined by the very process. It also determines the properties of functions of the observation as we have seen, for example, in the analysis of variance. We shall now go one step further and show how the randomization process, together with the analysis of variance, leads to a simple procedure for testing hypotheses about treatment effects. We shall give this development in some detail so that the reader can see immediately the application and extension to more complex designs to be discussed in later chapters.

6.5.1 Enumerating Randomizations

The outcome of our completely randomized assignment is that we have a plan associating treatments to EUs. We then perform the experiment and obtain a data table which may look like this, for example,

EU #	1	2	3	...	N
Treatment	3	1	4	...	7
Response	y_{32}	y_{13}	y_{41}	...	y_{74}

where y_{ij} refers to the response for the j th application of treatment i . We wish to consider the hypothesis that the treatments have no differential effects, that is, we would observe the same response on an EU regardless of which treatment has been used. The experimental plan we have used is a random one of

$$s = \frac{N!}{(r!)^t} = \frac{(tr)!}{(r!)^t} \quad (6.31)$$

possible plans (assignments). Let us name and index these plans by Π_γ ($\gamma = 1, 2, \dots, s$). Then if treatments were without differential effects, we would have obtained exactly the same result, that is, the same responses, for each Π_γ . We can thus visualize data sets for the s plans as given in Table 6.3. In sketching this table we have inserted treatment plans that could have been used (the numbering is arbitrary). One of these plans is, of course, the plan we have actually used, say Π_1 . The observations from the experiment using Π_1 are labeled z_1, z_2, \dots, z_n . Under the null hypothesis that there are no differences among the treatment effects the same observations would have been obtained from any of the other plans. For purposes of the analysis the observations would then have to be relabeled indicating for each z_ℓ ($\ell = 1, 2, \dots, N$) which treatment had been applied to the ℓ -th EU. For example for Π_1 we would have $z_1 = y_{21}$, $z_2 = y_{51}$, $z_3 = y_{31}$, \dots , $z_N = y_{1r}$ (assuming that the treatments are applied sequentially to the EU, starting with EU 1 and ending with EU N).

Table 6.3 Possible Outcomes for CRD

EU #	1	2	3	...	N
Response	z_1	z_2	z_3	...	z_N
Plan 1	2	5	3	...	1
2	1	2	2	...	4
3	3	1	4	...	2
\vdots	\vdots	\vdots	\vdots		\vdots
s	1	5	1	...	4

6.5.2 Randomization Test

We now construct a criterion that is determined by the treatment plan whether actual or potential. A criterion depends on the treatment plan if and only if when the observations are indexed by (ij) with i denoting the treatment and j the replication (application) within the treatment, it is invariant under permutations of j within i . The variety of such criteria is, of course, essentially unlimited. In order to exposit the idea we give some examples:

- (i) the sum of squares for treatments, $SS(T)$, in the ANOVA table,
- (ii) $MS(T)/MS(E)$,
- (iii) the range of treatment totals,
- (iv) the range of medians of the treatment groups (supposing there are no ties and an odd number of applications of each treatment),
- (v) the sum of squares (or range) of trimmed treatment means,
- (vi) the sum of squares (or range) of Winzorized treatment means,
- (vii) the sum of squares of robust estimates of treatment means.

Having chosen a criterion C , say, we are able to evaluate C for all possible plans Π_γ , giving us a set of numbers $\{C_\gamma, \gamma = 1, 2, \dots, s\}$. Actually, there are only $s^* = s/(t!)$ different values C_γ since each permutation of the treatment labels yields the same value for C . We denote the s^* different values by C_{γ^*} . Amongst these is the number associated with our actual plan, which we denote by C_a . Then we *declare* that the significance level (SL) against the null hypothesis of no differential treatment effects, with the chosen criterion, is

$$SL = \frac{1}{s} [\text{number of } C_\gamma (\gamma = 1, 2, \dots, s) \geq C_a] \quad (6.32a)$$

$$SL = \frac{1}{s^*} [\text{number of } C_{\gamma^*} (\gamma^* = 1, 2, \dots, s^*) \geq C_a] \quad (6.32b)$$

(Note that the number of C_γ on the right-hand side of (6.32a) and (6.32b) includes C_α , so that always $SL \geq 1/s^*$).

We then assert: Under the null hypothesis of no differential treatment effects if α is an achievable level, which will be a multiple of $1/s^*$, the probability of obtaining a significance level less than or equal to α is α . This assertion is obvious since under the null hypothesis we are observing with probability $1/s^*$ one of the numbers of the set $\{C_{\gamma^*}, \gamma^* = 1, 2, \dots, s^*\}$.

The procedure just described is called the *randomization test* (see Chapter 5). We illustrate it with the following example.

EXAMPLE 6.4: Suppose we have $t = 3$ treatments and $r = 3$ replications for each treatment. Let our plan, Π_1 say, and the observed responses be as follows:

EU	1	2	3	4	5	6	7	8	9
Trt	1	3	2	1	2	1	3	3	2
y	7.58	11.61	9.97	8.56	11.03	8.82	10.32	11.73	10.06

This plan is one out of $s = 9!/(3!)^3 = 1680$. Using $C = SS(T)$ as our test criterion we obtain $C_\alpha = SS(T)_1 = 13.2956$. This is one of $s^* = 1680/6 = 280$ different values for all possible values. It is not difficult to write a computer program to enumerate all possible 1680 plans and their associated 280 different $SS(T)$ -values. As a result we obtain for the significance level using (6.32a) or (6.32b)

$$SL = \frac{12}{1680} = \frac{2}{280} = .00714$$

Inspection of the plan Π_1 above suggests that in this case it is not necessary to spell out all different plans in order to obtain SL: it is clear that the largest $SS(T)$ -value, using the y -values above, is obtained from the plan

EU	1	2	3	4	5	6	7	8	9
Trt	1	3	2	1	3	1	2	3	2

which has Treatment 1 associated with the 3 lowest observations and Treatment 3 with the 3 highest observations, providing thus the largest mean separation among the treatments and hence the highest $SS(T)$, namely 14.8623. The plan above is obtained from the plan Π_1 by simply interchanging the treatments assigned to EUs 5 and 7 and thereby interchanging the third and fourth highest observations. As a consequence Π_1 then leads to the second largest mean separation and hence to the second largest $SS(T)$. Since each plan has five additional permutations with the same $SS(T)$ the result for SL as given above follows immediately. \square

Unfortunately, such simple arguments are not always possible. This means that usually all plans have to be enumerated and each C -value computed in order to obtain the significance level.

6.6 APPROXIMATING THE RANDOMIZATION TEST

We saw in the previous section that even for small t and r the number of possible plans s under randomization is quite large. Further evidence of the rapid increase in s for moderate values of t and r is given in Table 6.4.

These numbers indicate that even though it is possible today, in the computer age, to spin out all possible plans and proceed with the randomization analysis as described in Section 6.5, the procedure becomes rather cumbersome. It is in this context that we shall discuss an approximation to the randomization test by the F -test (see also Kempthorne, 1955) as suggested by the GMNLM theory (see Chapter 4).

6.6.1 Moments of the Test Statistic

We note first that under the null hypothesis, $H_0 : \tau_1 = \tau_2 = \dots \tau_t = 0$ (remember from Section 6.3.1 that $\sum \tau_i = 0$), using the results of Sections 6.2.2, 6.3.1, and 6.4.2

Table 6.4 Number of Experimental Plans

Treatments (t)	Number of Replications (r)	Plans (s)	Different C -Values (s^*)
2	4	70	35
	5	210	105
	6	924	462
	7	3,432	1,716
3	3	1,680	280
	4	34,650	5,775
	5	252,252	42,042
4	2	2,520	105
	3	369,600	15,400
	4	63,063,000	2,627,625
5	2	113,400	945
	3	168,168,000	1,401,400

$$\begin{aligned}
SS(T) + SS(E) &= SS(\text{Total}) \\
&= \sum_{ij} (y_{ij} - \bar{y}_{..})^2 \\
&= \sum_{ij} \omega_{ij}^2 \\
&= \sum_{ij} \left(\sum_k \delta_{ij}^k u_k \right)^2 \\
&= \sum_{ij} \sum_k \delta_{ij}^k u_k^2 + \sum_{ij} \sum_{k \neq k'} \delta_{ij}^k \delta_{ij}^{k'} u_k u_{k'} \\
&= \sum_k u_k^2
\end{aligned}$$

is a constant and, by definition, equal to $(N - 1)\sigma_u^2$. Instead of using $SS(T)$ as our test criterion we could, therefore, just as well consider

$$Z = \frac{SS(T)}{SS(T) + SS(E)}. \quad (6.33)$$

We know that under GMNLM theory the quantity

$$F = \frac{MS(T)}{MS(E)}$$

is distributed as $F_{t-1, t(r-1)}$ (see Chapter 4). It is then a fact that

$$\frac{(t-1)F}{t(r-1) + (t-1)F} = \frac{SS(T)}{SS(T) + SS(E)} = Z$$

follows a beta distribution with density

$$f(z)dz = \frac{1}{B\left(\frac{t-1}{2}, \frac{t(r-1)}{2}\right)} z^{\frac{t-1}{2}-1} (1-z)^{\frac{t(r-1)}{2}-1} dz$$

for $0 \leq z \leq 1$, that is, a beta (α, β) with $\alpha = (t-1)/2$, $\beta = t(r-1)/2$ (see Johnson, Kotz and Balakrishnan, 1995, p.327). From the properties of the beta distribution we know that if a random variable X is beta (α, β) , then

$$E(X^k) = \frac{B(\alpha + k, \beta)}{B(\alpha, \beta)} = \frac{(\alpha + k - 1)!}{(\alpha + \beta + k - 1)!} \cdot \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!}.$$

In particular, we have

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad (6.34)$$

and

$$E(X^2) = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)}. \quad (6.35)$$

We now consider the first two moments of Z , our test criterion, under GMNLM theory and randomization theory. Since Z , as defined in (6.33), is beta $[(t-1)/2, t(r-1)/2]$, it follows from (6.34) that

$$E(Z) = \frac{t-1}{tr-1} \quad (6.36)$$

and from (6.35) that

$$E(Z^2) = \frac{(t-1)(t+1)}{(tr-1)(tr+1)}. \quad (6.37)$$

We now consider $E_R(Z)$ and $E_R(Z^2)$. We have already shown (see Section 6.4) that, under the null hypothesis

$$E_R[\text{SS}(T)] = (t-1)\sigma_u^2.$$

It follows then that

$$E_R(Z) = \frac{t-1}{tr-1}. \quad (6.38)$$

To work out $E_R[\text{SS}(T)]^2$ and hence $E_R(Z^2)$ is rather tedious and lengthy. We shall not give a derivation but use a result from Richards (1980) which yields

$$\begin{aligned} E_R[\text{SS}(T)]^2 &= \left\{ \frac{2(t-1)(r-1)(N^2 - 3N + 3)}{r(N-1)^2(N-2)(N-3)} + \frac{(N-r)^2}{r^2(n-1)^2} \right\} \left(\sum_k u_k^2 \right)^2 \\ &\quad - \frac{2N(r-1)(t-1)}{r(N-1)(N-2)(N-3)} \sum_k u_k^4. \end{aligned} \quad (6.39)$$

For large r , in the sense that $1/r$ and hence $1/N$ are small compared to 1, (6.39) can be approximated by

$$\begin{aligned} E_R[\text{SS}(T)]^2 &\cong \frac{2(t-1) + (t-1)^2}{t^2 r^2} \left[\sum u_k^2 \right]^2 \\ &= \frac{(t-1)(t+1)}{t^2 r^2} \left[\sum u_k^2 \right]^2 \\ &= \frac{(t-1)(t+1)}{t^2 r^2} (N-1)^2 \sigma_u^4. \end{aligned}$$

It follows then that

$$E_R(Z^2) \cong \frac{(t-1)(t+1)}{t^2 r^2}. \quad (6.40)$$

6.6.2 Approximation by the F -Test

Comparing (6.38) with (6.36) and (6.40) with (6.37) we can say that the distributions of Z under normal theory and under randomization theory are in good agreement with respect to the first two moments. We take this to mean that the randomization distribution of Z is “fairly accurately” represented by the beta distribution. This implies that the randomization distribution of $MS(T)/MS(E)$ is fairly accurately represented by the F -distribution with $t - 1$ and $t(r - 1)$ d.f. It is in this sense that we consider the ordinary F -test for testing the hypothesis of no differences among the treatment effects as a good approximation to the randomization test discussed in Section 6.5. We mention here that these ideas and results go back to Fisher (1935), Pitman (1937), and Welch (1937).

It should be pointed out that the preceding discussion was based entirely on the derived linear model based on the assumption of additivity in the strict sense. As discussed earlier (see Section 6.3) for practical applications the model obtained under the assumption of additivity in the broad sense is more realistic. The question then arises: How should we test the hypothesis of no treatment differences under this model? There does not seem to be an easy way, if any at all, to derive a result for model (6.22) analogous to the one just derived for model (6.8). Nevertheless, we take the results of this section as a strong indication that the usual F -test as suggested by the ANOVA is an appropriate test procedure.

6.6.3 Simulation Study

The arguments given above for suggesting that the randomization test can be approximated by the usual F -test is of course, not entirely satisfactory. The following questions remain: (i) What happens for small r , and (ii) to what extent does agreement of the first two moments imply agreement of both distributions? Although we cannot provide an analytical solution, we can give some indication that, indeed, the agreement between both distributions is quite good in general.

We shall illustrate the argument in terms of a simple example. Suppose we have $t = 4, r = 8$. The total number of randomizations is $s = 2.39 \times 10^{24}$. Except for very powerful computers it is a nearly impossible task to enumerate all possible randomizations as described in Section 6.5, hence approximation of the randomization test by the F -test appears to be the only practical solution. To demonstrate that this is, indeed, a reasonable approach we conduct the following simulation experiment. Out of all possible randomizations we select at random $s' = 500$ randomizations of the 4 treatments to the 32 EUs. Assigning the (arbitrarily chosen) responses 0, 1, 2, 3, to 8 EUs each we then compute for each of the s' arrangements the quantity $F = MS(T)/MS(E)$, denoted by $F^{(1)}, F^{(2)}, \dots, F^{(500)}$. For each $F^{(i)} (i = 1, 2, \dots, 500)$ we then obtain the significance level in two ways:

- (i) based on the F -distribution with 3 and 28 d.f.,
- (ii) based on the rank among all 500 F -values as explained in (6.32a).

We denote these significance levels by NSL and RSL, respectively. A plot of RSL vs. NSL is given in Figure 6.1. It shows that both significance levels are in “close”

Table 6.5 Comparison of Significance Levels

Rank	F -Value*	NSL	RSL
1	8.65863	0.00032	0.002
2	6.55319	0.00170	0.004
3	4.88889(2)	0.00740	0.008
5	4.62305(2)	0.00949	0.012
7	4.49383(4)	0.01073	0.020
11	4.36697	0.01211	0.022
12	4.12012(3)	0.01537	0.028
16	3.65217	0.02436	0.032
17	3.43020(2)	0.03045	0.036
19	3.32203	0.03399	0.038
20	3.21569	0.03788	0.040
21	3.00826	0.04691	0.042
22	2.80759(6)	0.05781	0.054
28	2.70968(2)	0.06406	0.058
30	2.61333(6)	0.07091	0.070

*Values in parentheses indicate frequency of occurrence.

agreement, but the line with slope 1 through the points indicates that RSL is (with some exceptions) always slightly larger than NSL. Some indication of the discrepancy for small significance levels is given in Table 6.5. Some of the “large” discrepancies are, of course, due to the discreteness of RSL and the fact that the same F -value may occur more than once. From a practical point of view, however, the agreement between NSL and RSL is quite remarkable.

In the discussion above, the reader should keep in mind that this is only one example intended to illustrate a point, namely to support the theoretical result that, under certain conditions, the randomization distribution can be approximated by the F -distribution. This is, of course, only the beginning of what would have to be an extensive Monte Carlo study, using different values for t and r and for the responses y . Suffice it to say here that for a number of different values similar results were obtained giving plausible credence to the validity of our assertion (see also Kempthorne and Doerfler, 1969). Hence, from now on we shall use the F -test as an approximation to the randomization test for testing the hypothesis $H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$.

We conclude this discussion by pointing out that in most cases the null hypothesis of the equality of treatment effects is not the most important hypothesis to test. For a

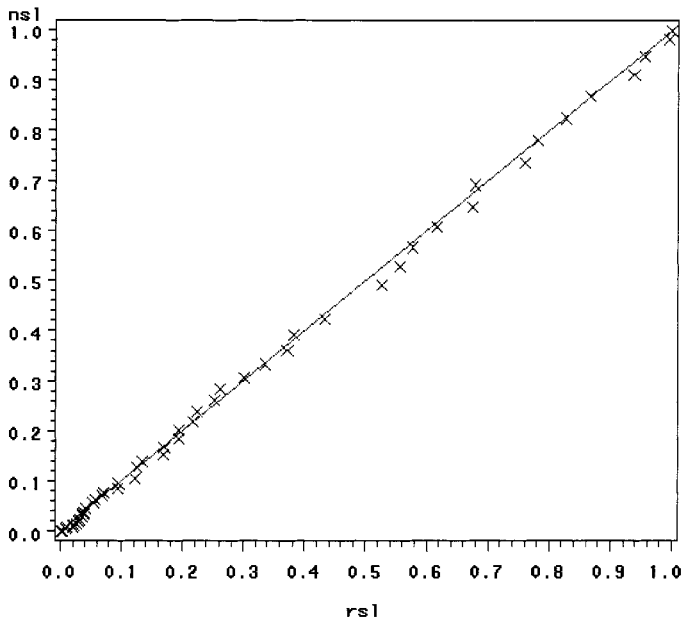


Figure 6.1 Plot of Significance Levels of Randomization Test vs. Approximating *F*-Test.

well designed interested experiment the researcher is often more interested in specific treatment comparisons. Such questions will be discussed in more detail in Chapter 7. This does not diminish the importance of the ANOVA as a data analysis technique, as one of its useful functions is the estimation of the error variance component σ_e^2 , or as shown in Section 6.9 the error variance components σ_ε^2 and σ_η^2 .

6.7 CRD WITH UNEQUAL NUMBERS OF REPLICATIONS

Although in most applications of the CRD each treatment will be replicated the same number of times, namely r , it is not uncommon to have unequal numbers of replications, say r_i for the i th treatment ($i = 1, 2, \dots, t$). We may have, for example, the situation that one treatment, say treatment 1, is the control or standard with which we wish to compare the other treatments [see also Section 7.5.7]. It seems then reasonable to obtain especially good information about treatment 1, that is, to have more replications for treatment 1 than for the other treatments. Or, it may be that among the t treatments some are more important than others, suggesting different numbers of replications for the two sets of treatments. Another reason for having unequal r_i 's may simply be the fact that the observations from some EUs may be missing (independent of the treatments).

6.7.1 Randomization

Just as for the case of the equireplicate CRD, the randomization procedure can be performed by using random numbers (see Section 6.2.1). From a practical point of view, the process can be implemented by using appropriate statistical software, e.g. SAS PROC PLAN (SAS Institute, Inc., 2002–2003). We illustrate this with the following example.

EXAMPLE 6.5: Suppose we consider a CRD with $t = 4$, $r_1 = 4$, $r_2 = r_3 = r_4 = 2$. The SAS input statements and the output are given in Table 6.6. \square

6.7.2 The Model and ANOVA

It is obviously much more complicated to derive the randomization analysis for the situations described above. The basic model, however, is still model (6.22),

$$y_{ij} = \mu + \tau_i + e_{ij}$$

with $i = 1, 2, \dots, t$; $j = 1, 2, \dots, r_i$. The ANOVA table for equal numbers of replications (Table 6.2) is modified easily to accommodate unequal numbers of replications and is given in Table 6.7.

Just as for the equal number case (Section 6.6) it can be illustrated through Monte Carlo studies that the randomization test for testing $H_0 : \tau_1 = \tau_2 = \dots = \tau_t$ can be approximated by the F -test

$$F = \frac{MS(T)}{MS(E)}$$

with $t - 1$ and $(\sum r_i - t)$ d.f.

6.7.3 Comparing Randomization Test and F -Test

To illustrate the general agreement between the significance levels for both tests, that is, RSL and NSL, we give below the result for a simulation run for $t = 4$ and $r_1 = r_2 = 4$, $r_3 = r_4 = 2$. In Figure 6.2 we show the relationship between RSL and NSL based on a random sample of 1000 randomizations, using the same procedure as described in Section 6.6. Obviously, a much broader simulation study would have to be done to give more support to our claim that for the unequal replication CRD the randomization test can be approximated by the F -test, but a small number of simulations have led to results similar to those given in Figure 6.2. We found, in general, a good agreement between RSL and NSL for random samples of 1,000 randomizations as illustrated in Figure 6.2. Note also that these results are similar to those given in Figure 6.1. Obviously when the experiment is small the approximation may not be good, but then the randomization test can be done easily on a computer.

6.8 NUMBER OF REPLICATIONS

One question that is being asked often, and it is an important question, is: How many replications are needed for each treatment? The reason for asking this question (which

Table 6.6 Randomization Procedure for CRD with Unequal Replications

a.) Input Statements:

```
proc plan seed=13396;
factors unit=10;
treatments treat=10 cyclic (1 1 1 1 2 2 2 3 3 4 4);
output out=CRD;
title1 'COMPLETELY RANDOMIZED DESIGN';
title2 'WITH UNEQUAL REPLICATIONS';
title3 '(t=4, r1=4 r2=r3=r4=2, N=10)';
run;

proc sort out=CRD;
by unit;
run;

proc print;
run;
```

b.) Output:

COMPLETELY RANDOMIZED DESIGN
WITH UNEQUAL REPLICATIONS
(t=4, r1=4 r2=r3=r4=2, N=10)

The PLAN Procedure

Plot Factors

Factor	Select	Levels	Order
unit	10	10	Random

Treatment Factors

Factor	Select	Levels	Order	Initial Block / Increment
treat	10	10	Cyclic	(1 1 1 1 2 2 2 3 3 4 4) / 1

-----unit-----										-----treat-----									
8	4	3	10	9	1	5	7	6	2	1	1	1	1	2	2	3	3	4	4

Table 6.6 (Continued)

	Obs	unit	treat
	1	1	2
	2	2	4
	3	3	1
	4	4	1
	5	5	3
	6	6	4
	7	7	3
	8	8	1
	9	9	2
	10	10	1

Table 6.7 ANOVA for CRD with Unequal Numbers of Replications

Source	d.f.	SS	MS	$E(\text{MS})$
Treatments	$t - 1$	$\sum_i r_i (\bar{y}_{i.} - \bar{y}_{..})^2$	$\text{MS}(T)$	$\sigma_e^2 + \frac{1}{t - 1} \sum_i r_i \tau_i^2$
Error	$\sum_i r_i - t$	$\sum_{i,j} (y_{ij} - \bar{y}_{i.})^2$	$\text{MS}(E)$	σ_e^2
Total	$\sum_i r_i - 1$	$\sum_{i,j} (y_{ij} - \bar{y}_{..})^2$		

is often referred to – incorrectly – as a question of “sample size”) is to “assure” that the experiment is sensitive enough to detect differences among the treatments if there are any, that is, to reject the null hypothesis of no treatment differences in the ANOVA F -test. As it stands, however, the question cannot be answered without further input specific to the particular investigation at hand.

6.8.1 Power of the F -Test

Based on our discussion above we shall use the normal (Gaussian) independent error model and the associated central and noncentral F -distributions to examine this question, using the notion of the power of the F -test. More specifically, the sensitivity or power of the F -test, denoted by $1 - \beta$, where β is the probability of a Type II error, depends on

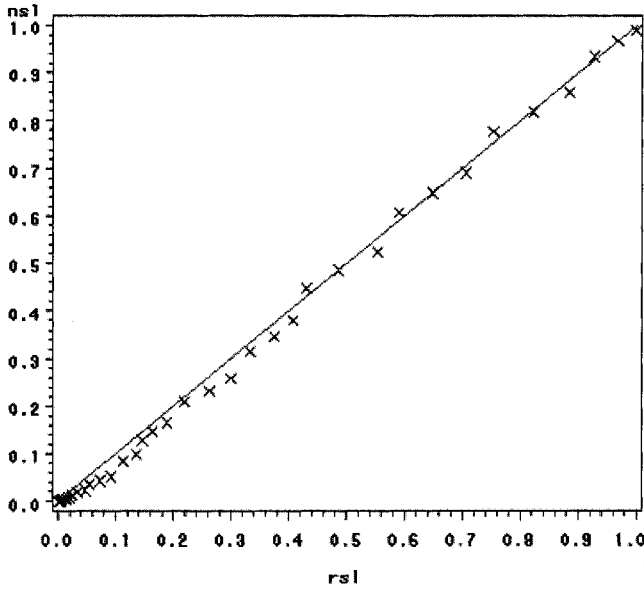


Figure 6.2 Plot of Significance Levels of Randomization Test (rsl) versus Approximating F -Test (nsl).

- (i) the size of the test, that is, probability of the Type I error, α ;
- (ii) the degrees of freedom, $t - 1$ and $t(r - 1)$;
- (iii) the noncentrality parameter

$$\lambda = \frac{r \sum_i \tau_i^2}{2\sigma_e^2} \quad (6.41)$$

of the noncentral F -distribution, where the τ_i are the true values of the treatment effects as specified under the alternative hypothesis. The general procedure then is to specify α , $1 - \beta$, and λ/r and ask: How many replications, r , are needed to detect, with probability $1 - \beta$, treatment differences as specified by λ/r if we use a test of size α ?

It is, of course, not difficult to specify α and β . We usually take $\alpha = .05$ or $\alpha = .10$ as those seem to be reasonable values for the risk of committing a Type I error, that is, concluding that there are differences among the treatments, when in fact there are none. A bit more difficult is the choice of β or rather $1 - \beta$, the probability for concluding that there are differences among the treatments when they indeed exist. From a practical point a reasonable choice is $1 - \beta = .80$ although this choice, as all the others, depends on the particular problem under consideration. By far the most difficult choice is that of λ/r , because that, after all, represents the true state of nature, something we do not know. How does one get out of that dilemma?

6.8.2 Smallest Detectable Difference

It is at this point that the subject matter knowledge of the investigator becomes very important. Since we do not know the true state of nature, we might ask: What minimum difference between the two extreme treatments, the best and the worst, is worth detecting with probability of at least $1 - \beta$, if such a difference exists? This question is best visualized for the simple case $t = 2$. Suppose we want to compare two drugs, an established drug A versus an experimental drug B expected to be better than A . How much greater has the therapeutic effect of B have to be before it is worth further development and marketing? There is obviously a minimum difference in therapeutic effects before B is worth developing, both from a medical as well as financial point of view.

For the general case let us denote the difference between the largest treatment effect, τ_{\max} , and the smallest treatment effect, τ_{\min} , by

$$\Delta = \tau_{\max} - \tau_{\min}. \quad (6.42)$$

For any set of τ_i ($i = 1, 2, \dots, t$) satisfying (6.42) the smallest λ/r is obtained when the remaining $t - 2$ treatment effects τ_i are equal to $(\tau_{\max} + \tau_{\min})/2$. Since the τ_i are defined such that $\Sigma \tau_i = 0$, this means that

$$\tau_{\max} = \frac{\Delta}{2}, \quad \tau_{\min} = -\frac{\Delta}{2}, \quad \tau_i = 0 \quad \text{otherwise.} \quad (6.43)$$

It then follows from (6.41) that

$$\lambda/r = \frac{2 \frac{\Delta^2}{4}}{2\sigma_e^2} = \frac{\Delta^2}{4\sigma_e^2}. \quad (6.44)$$

It is well known that the power of the F -test is an increasing function of λ . Hence the power of the F -test with λ given by (6.44) has the smallest value for all situations subject to (6.42). Tables and charts for the power of the F -test in terms of λ (or suitable functions of λ) and d.f. $\nu_1 = t - 1$ and $\nu_2 = t(r - 1)$ are available (Tang, 1938; Pearson and Hartley, 1970; Odeh and Fox, 1975) and can be used to obtain iteratively a suitable value for r , given α , $1 - \beta$, Δ , σ_e^2 . For a description of this procedure we refer the reader to Scheffé (1959).

A more convenient set of tables was developed specifically for the determination of the number of replications by Bowman and Kastenbaum (1975). They follow essentially the same arguments as given above but in terms of

$$\Delta^* = \frac{\tau_{\max} - \tau_{\min}}{\sigma_e} = \frac{\tau_{\max} - \tau_{\min}}{\sqrt{\sigma_e^2 + \sigma_\eta^2}} \quad (6.45)$$

the standardized minimum difference between the two extreme treatment effects (this is sometimes referred to as the *effect size*). It is quite often easier to specify Δ^* rather than Δ , and specifying Δ^* absolves one from also specifying σ_e^2 . Selected parts of the tables from Bowman and Kastenbaum (1975) are reproduced in Table 6.8 for $t = 2, 3, 4, 5, 6, 8, 9, 10, 11, 13, 15, 20, 25, 30$, and $\alpha = .05$, $1 - \beta = .7, .8, .9$, and $r = 2, 3, \dots, 25$.

6.8.3 Practical Considerations

To show the use of these tables and the sometimes surprising results concerning the magnitude of r , we consider the following example.

EXAMPLE 6.6: Suppose we have $t = 3$, $\alpha = .05$. For different values of $1 - \beta = .7$, $.8$, $.9$ and $\Delta^* = 1.0, 1.5, 2.0$, we obtain from Table 6.8 the following values for r given in Table 6.9 (since the exact values for Δ^* are not represented in Table 6.8 we choose the next smaller value for Δ^* that is represented in Table 6.8). \square

Table 6.9 illustrates a number of points:

- (i) In some cases the number of replications seems surprisingly large. This often leads to disappointment when an experiment has been carried out without prior consideration of the number of replications. Generally too few replications are chosen resulting in a low power of the F -test which means that much effort may have been wasted.
- (ii) To detect small differences requires relatively more replications, in fact it may require more replications than is practical. It is therefore important to arrive at a realistic value for Δ^* .
- (iii) As the probability $1 - \beta$ of detecting an existing difference increases so does r , and in certain situations at an appreciable rate.

An alternative to using Table 6.8 is to use the Power Procedure in SAS (SAS Institute, Inc., 2002-2003). To use this procedure, rather than specifying Δ^* of (6.45) we need to specify the τ_i or $\mu + \tau_i$ ($i = 1, 2, \dots, t$) under the alternative hypothesis (referred to as group means in the procedure) and the standard deviation σ_e . In order to obtain the same results as described above in connection with Table 6.8, we specify, for given Δ^* , the τ_i as in (6.43) with $\Delta = \Delta^*$ and $\sigma_e = 1$. We illustrate this procedure in the following example.

EXAMPLE 6.6 (continued): Using $t = 3$, $\alpha = .05$, $\beta = .9$, and $\Delta^* = 1$, the SAS PROC POWER input is given in Table 6.10a and the output in Table 6.10b. The result is $r = 27$ (labeled N Per Group in the output), complementing the value (> 25) in Table 6.9. \square

The conclusion one should draw from this discussion is that it is important to present the investigator with a table like Table 6.9 prior to the experiment to explain the options available. It is not so important to decide whether one needs 20 or 21 replications, but rather that one needs about 20 and not 10 replications, for example.

If the investigator is more comfortable assigning a value to Δ rather than Δ^* , we need, of course, some information about σ_e^2 to determine r . Sometimes this information is available from previous similar experiments. In other cases one may have to do a preliminary study to estimate σ_e^2 . This estimate may be the $MS(E)$ from the ANOVA table for the preliminary study or simply an estimate of the variance using just one

Table 6.8 Values of Δ^* to Determine Numbers of Replications CRD*

<i>r</i>	$\frac{t = 2}{1 - \beta}$			$\frac{t = 3}{1 - \beta}$			$\frac{t = 4}{1 - \beta}$		
	.7	.8	.9	.7	.8	.9	.7	.8	.9
2	4.863	5.653	6.796	4.883	5.570	6.548	4.872	5.504	6.395
3	2.703	3.071	3.589	2.957	3.325	3.838	3.094	3.460	3.967
4	2.104	2.381	2.767	2.335	2.618	3.010	2.468	2.754	3.148
5	1.792	2.024	2.348	1.997	2.236	2.568	2.119	2.362	2.698
6	1.590	1.796	2.081	1.775	1.987	2.280	1.888	2.104	2.401
7	1.446	1.632	1.890	1.615	1.808	2.073	1.719	1.916	2.186
8	1.335	1.507	1.745	1.492	1.670	1.915	1.590	1.771	2.020
9	1.247	1.407	1.629	1.394	1.560	1.788	1.486	1.655	1.888
10	1.175	1.325	1.534	1.313	1.469	1.684	1.400	1.559	1.778
11	1.113	1.256	1.454	1.245	1.393	1.596	1.328	1.479	1.686
12	1.061	1.197	1.385	1.186	1.327	1.521	1.266	1.409	1.607
13	1.016	1.145	1.326	1.135	1.270	1.456	1.211	1.349	1.538
14	0.975	1.100	1.273	1.090	1.220	1.398	1.164	1.296	1.478
15	0.940	1.060	1.226	1.050	1.175	1.347	1.121	1.249	1.424
16	0.908	1.024	1.185	1.015	1.135	1.301	1.083	1.206	1.375
17	0.879	0.991	1.147	0.982	1.099	1.259	1.049	1.168	1.331
18	0.852	0.961	1.112	0.953	1.066	1.222	1.017	1.133	1.292
19	0.828	0.934	1.081	0.926	1.036	1.187	0.988	1.101	1.255
20	0.806	0.909	1.052	0.901	1.008	1.155	0.962	1.071	1.222
21	0.786	0.886	1.025	0.878	0.982	1.126	0.938	1.044	1.191
22	0.767	0.865	1.000	0.857	0.959	1.099	0.915	1.019	1.162
23	0.749	0.845	0.977	0.837	0.936	1.073	0.894	0.996	1.135
24	0.733	0.826	0.956	0.819	0.916	1.050	0.874	0.974	1.110
25	0.717	0.809	0.936	0.802	0.897	1.028	0.856	0.953	1.087

Table 6.8 (Continued)

<i>r</i>	$\frac{t = 5}{1 - \beta}$			$\frac{t = 6}{1 - \beta}$			$\frac{t = 7}{1 - \beta}$		
	.7	.8	.9	.7	.8	.9	.7	.8	.9
2	4.889	5.490	6.333	4.922	5.505	6.317	4.963	5.534	6.327
3	3.197	3.562	4.065	3.283	3.647	4.149	3.358	3.723	4.224
4	2.568	2.856	3.251	2.650	2.940	3.337	2.721	3.013	3.412
5	2.211	2.457	2.795	2.287	2.535	2.876	2.352	2.602	2.945
6	1.973	2.191	2.492	2.042	2.264	2.567	2.102	2.326	2.632
7	1.798	1.997	2.271	1.863	2.065	2.341	1.919	2.123	2.401
8	1.664	1.848	2.100	1.725	1.911	2.166	1.777	1.965	2.223
9	1.556	1.728	1.963	1.613	1.787	2.026	1.662	1.839	2.080
10	1.466	1.628	1.850	1.521	1.685	1.910	1.568	1.734	1.961
11	1.391	1.544	1.755	1.443	1.599	1.812	1.488	1.645	1.861
12	1.326	1.472	1.673	1.376	1.524	1.727	1.419	1.569	1.774
13	1.269	1.409	1.602	1.317	1.459	1.654	1.358	1.502	1.699
14	1.220	1.354	1.539	1.266	1.402	1.589	1.305	1.444	1.633
15	1.175	1.305	1.483	1.220	1.351	1.531	1.258	1.391	1.573
16	1.135	1.261	1.432	1.178	1.306	1.479	1.216	1.344	1.520
17	1.099	1.221	1.387	1.141	1.264	1.433	1.177	1.302	1.472
18	1.066	1.184	1.345	1.107	1.226	1.390	1.142	1.263	1.428
19	1.036	1.151	1.307	1.076	1.192	1.351	1.110	1.228	1.388
20	1.009	1.120	1.273	1.047	1.160	1.315	1.081	1.195	1.351
21	0.983	1.092	1.240	1.021	1.131	1.282	1.053	1.165	1.317
22	0.960	1.065	1.210	0.996	1.104	1.251	1.028	1.137	1.285
23	0.938	1.041	1.183	0.973	1.078	1.222	1.004	1.111	1.256
24	0.917	1.018	1.157	0.952	1.055	1.195	0.982	1.086	1.228
25	0.898	0.997	1.132	0.932	1.033	1.170	0.962	1.064	1.203

Table 6.8 (Continued)

r	$\frac{t=8}{1-\beta}$			$\frac{t=9}{1-\beta}$			$\frac{t=10}{1-\beta}$		
	.7	.8	.9	.7	.8	.9	.7	.8	.9
2	5.009	5.572	6.350	5.056	5.613	6.382	5.104	5.657	6.419
3	3.426	3.791	4.293	3.488	3.854	4.356	3.545	3.913	4.416
4	2.784	3.078	3.479	2.841	3.136	3.540	2.893	3.191	3.596
5	2.409	2.662	3.008	2.461	2.716	3.064	2.509	2.766	3.116
6	2.155	2.381	2.689	2.203	2.431	2.741	2.247	2.477	2.789
7	1.968	2.174	2.455	2.013	2.221	2.504	2.054	2.263	2.548
8	1.823	2.014	2.274	1.865	2.057	2.319	1.903	2.097	2.361
9	1.706	1.884	2.128	1.746	1.926	2.171	1.782	1.963	2.210
10	1.609	1.777	2.006	1.647	1.816	2.048	1.681	1.852	2.085
11	1.527	1.687	1.904	1.563	1.724	1.943	1.596	1.758	1.979
12	1.457	1.609	1.816	1.491	1.644	1.853	1.522	1.677	1.888
13	1.395	1.540	1.739	1.428	1.575	1.775	1.458	1.606	1.808
14	1.340	1.480	1.671	1.372	1.513	1.706	1.401	1.544	1.738
15	1.292	1.427	1.611	1.323	1.459	1.644	1.351	1.488	1.675
16	1.248	1.379	1.556	1.278	1.410	1.589	1.305	1.438	1.619
17	1.209	1.335	1.507	1.238	1.365	1.539	1.264	1.393	1.568
18	1.173	1.295	1.462	1.201	1.325	1.493	1.227	1.351	1.521
19	1.140	1.259	1.421	1.167	1.288	1.451	1.192	1.314	1.479
20	1.110	1.226	1.384	1.136	1.253	1.413	1.161	1.279	1.440
21	1.082	1.195	1.349	1.108	1.222	1.377	1.131	1.247	1.403
22	1.056	1.166	1.316	1.081	1.193	1.344	1.104	1.217	1.370
23	1.032	1.139	1.286	1.057	1.165	1.313	1.079	1.189	1.338
24	1.009	1.114	1.258	1.033	1.140	1.285	1.056	1.163	1.309
25	0.988	1.091	1.232	1.012	1.116	1.258	1.033	1.139	1.282

Table 6.8 (Continued)

	$\frac{t=11}{1-\beta}$			$\frac{t=13}{1-\beta}$			$\frac{t=15}{1-\beta}$		
r	.7	.8	.9	.7	.8	.9	.7	.8	.9
2	5.152	5.702	6.458	5.245	5.792	6.541	5.334	5.879	6.625
3	3.599	3.968	4.472	3.697	4.069	4.576	3.785	4.161	4.670
4	2.942	3.241	3.649	3.030	3.333	3.744	3.109	3.415	3.830
5	2.553	2.812	3.164	2.633	2.895	3.251	2.705	2.970	3.329
6	2.288	2.519	2.834	2.361	2.596	2.914	2.426	2.664	2.986
7	2.091	2.303	2.590	2.160	2.374	2.665	2.220	2.437	2.732
8	1.939	2.134	2.400	2.002	2.201	2.470	2.059	2.260	2.533
9	1.815	1.998	2.247	1.875	2.061	2.313	1.929	2.117	2.372
10	1.713	1.885	2.120	1.770	1.945	2.183	1.820	1.998	2.239
11	1.626	1.790	2.012	1.680	1.847	2.073	1.728	1.897	2.126
12	1.551	1.707	1.920	1.603	1.762	1.977	1.649	1.810	2.029
13	1.486	1.635	1.839	1.536	1.688	1.894	1.580	1.734	1.944
14	1.428	1.572	1.767	1.476	1.622	1.821	1.519	1.667	1.868
15	1.376	1.515	1.704	1.423	1.564	1.755	1.464	1.607	1.801
16	1.330	1.464	1.646	1.375	1.512	1.696	1.415	1.554	1.741
17	1.288	1.418	1.595	1.332	1.464	1.643	1.371	1.505	1.686
18	1.250	1.376	1.547	1.293	1.421	1.594	1.330	1.460	1.636
19	1.215	1.338	1.504	1.257	1.381	1.550	1.293	1.420	1.591
20	1.183	1.302	1.464	1.223	1.345	1.509	1.259	1.382	1.549
21	1.153	1.270	1.427	1.193	1.311	1.471	1.228	1.348	1.510
22	1.126	1.239	1.393	1.164	1.279	1.436	1.198	1.315	1.474
23	1.100	1.211	1.361	1.138	1.250	1.403	1.171	1.285	1.440
24	1.076	1.184	1.332	1.113	1.223	1.373	1.145	1.257	1.409
25	1.053	1.160	1.304	1.090	1.197	1.344	1.122	1.231	1.379

Table 6.8 (Continued)

r	$\frac{t = 20}{1 - \beta}$			$\frac{t = 25}{1 - \beta}$			$\frac{t = 30}{1 - \beta}$		
	.7	.8	.9	.7	.8	.9	.7	.8	.9
2	5.539	6.086	6.829	5.722	6.272	7.018	5.886	6.441	7.191
3	3.977	4.359	4.877	4.138	4.527	5.053	4.279	4.674	5.208
4	3.278	3.592	4.015	3.419	3.739	4.171	3.542	3.868	4.307
5	2.856	3.129	3.497	2.983	3.261	3.637	3.092	3.376	3.758
6	2.565	2.810	3.139	2.681	2.931	3.268	2.780	3.036	3.379
7	2.349	2.572	2.874	2.455	2.684	2.993	2.548	2.781	3.095
8	2.179	2.386	2.666	2.279	2.491	2.777	2.365	2.582	2.874
9	2.042	2.236	2.498	2.136	2.335	2.603	2.217	2.420	2.694
10	1.928	2.111	2.359	2.017	2.205	2.458	2.094	2.286	2.544
11	1.831	2.005	2.240	1.916	2.094	2.335	1.989	2.171	2.417
12	1.747	1.913	2.138	1.829	1.999	2.228	1.899	2.073	2.307
13	1.674	1.833	2.048	1.752	1.916	2.135	1.820	1.986	2.211
14	1.610	1.763	1.969	1.685	1.842	2.053	1.750	1.910	2.126
15	1.552	1.700	1.899	1.625	1.776	1.980	1.687	1.842	2.050
16	1.500	1.643	1.835	1.571	1.717	1.914	1.631	1.781	1.981
17	1.453	1.591	1.778	1.521	1.663	1.854	1.580	1.725	1.920
18	1.410	1.544	1.725	1.477	1.614	1.799	1.534	1.674	1.863
19	1.371	1.502	1.677	1.436	1.569	1.749	1.491	1.628	1.811
20	1.335	1.462	1.633	1.398	1.528	1.703	1.452	1.585	1.764
21	1.302	1.425	1.592	1.363	1.490	1.661	1.416	1.545	1.720
22	1.271	1.391	1.554	1.331	1.454	1.621	1.382	1.509	1.679
23	1.242	1.360	1.519	1.300	1.421	1.584	1.351	1.474	1.641
24	1.215	1.330	1.486	1.272	1.390	1.550	1.321	1.442	1.605
25	1.189	1.302	1.455	1.246	1.361	1.518	1.294	1.412	1.572

*Reproduced from K. O. Bowman and M. A. Kastenbaum, "Sample size requirement: Single and double classification experiments" in *Selected Tables in Mathematical Statistics*, Vol. 3 (1975), by permission from the authors and the American Mathematical Society.

Table 6.9 Number of Replications in CRD

$1 - \beta$	Δ^*		
	1.0	1.5	2.0
.7	17	8	5
.8	21	10	6
.9	> 25	13	8

treatment, for example the control treatment. Although preliminary studies are usually rather small, they should be sufficiently large to get a reliable estimate of σ_e^2 , that is, an estimate based on a sufficient number of degrees of freedom.

6.9 SUBSAMPLING IN A CRD

As we have pointed out earlier (see also Section 2.3), a careful distinction must be made between experimental units (EU) and observational (sampling) units (OU). Until now we have considered in this chapter the situation where EUs and OUs are identical. One consequence of this situation is that even though in the formulation of a linear model for observations from a CRD we distinguish between experimental error (ε_{ij}) and observational error (η_{ijk}), we cannot separate the two error terms in the analysis and hence we combine them usually into one error term (e_{ij}). There are, however, situations where EUs and OUs are not identical. We refer to the example in Section 2.3, where a class of students is the EU and the individual students are the OUs. This situation is generally referred to as a *CRD with subsampling*.

6.9.1 Subsampling Model

Suppose then we have t treatments, each replicated r' times, and each EU has n OUs, that is, n observations are obtained from each EU. An extension of model (6.19) can then be written as

$$y_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \eta_{ijk} \quad (6.46)$$

($i = 1, 2, \dots, t; j = 1, 2, \dots, r'; k = 1, 2, \dots, n$) where ε_{ij} represents the experimental error and η_{ijk} the observational error. According to our convention, we treat the ε_{ij} as i.i.d. $(0, \sigma_\varepsilon^2)$ and the η_{ijk} as i.i.d. $(0, \sigma_\eta^2)$. We note that

$$\text{var}(y_{ijk}) = \sigma_\varepsilon^2 + \sigma_\eta^2 = \sigma_e^2$$

just as before except that we can now separate the two variance components, or rather their estimates. This becomes obvious from the ANOVA table (see Table 6.11) associated with the linear model (6.46) which in linear model theory is referred to as a

Table 6.10 Determination of Number of Replications

a.) Input Statements:

```
proc power;
onewayanova test=overall
groupmeans = -.5| 0| .5
stddev = 1
npergroup = .
power = .9
alpha = .05;
run;
```

b.) Output:

The SAS System

The POWER Procedure Overall F Test for One-Way ANOVA

Fixed Scenario Elements

Method	Exact
Alpha	0.05
Group Means	-0.5 0 0.5
Standard Deviation	1
Nominal Power	0.9

Computed N Per Group

Actual Power	N Per Group
0.908	27

two-fold nested classification: the EUs are nested within the treatments and the OUs are nested within the EUs (see also Section 4.12 for the definition of a nested classification). The ANOVA table can be obtained easily from the following identity

$$y_{ijk} = \bar{y} \dots + (\bar{y}_{i..} - \bar{y} \dots) + (\bar{y}_{ij.} - \bar{y}_{i..}) + (\bar{y}_{ijk} - \bar{y}_{ij.})$$

mimicking model (6.46) and proceeding along the same lines as indicated in Section 6.4.

6.9.2 Inferences with Subsampling

It follows from Table 6.11 that in order to test the null hypothesis of no treatment differences we use the F -test (again, as an approximation to the randomization test)

$$F = \frac{MS(T)}{MS(EE)} \quad (6.47)$$

with $t - 1$ and $t(r' - 1)$ d.f. Also, as pointed out earlier, and as is obvious from the $E(MS)$ in Table 6.11, the experimental and observational error variance components can be estimated separately, namely

$$\hat{\sigma}_\eta^2 = MS(OE) \quad (6.48)$$

and

$$\hat{\sigma}_\varepsilon^2 = [MS(EE) - MS(OE)]/n. \quad (6.49)$$

Since the use of several observations per EU, that is, subsampling, does not constitute replication of treatments, and since the d.f. for the F -test (6.47) are determined by t and r' and not by n , we may ask: What are the benefits that arise from subsampling? We have already pointed out one benefit, namely the separation of the estimates for experimental and observational error variance components. This allows us a closer look at our experimental and measurement techniques or rather the quality of these techniques expressed in terms of their variability. If we find, for example, that $\hat{\sigma}_\eta^2$ is unreasonably large, we may try to improve the reliability of our measurement process; or if $\hat{\sigma}_\varepsilon^2$ is quite large, we may take another look at the EUs and their “homogeneity” and decide that we could reduce the experimental error by using supplementary information (see Chapter 8) or another design such as a randomized complete block design (see Chapter 9). Reduction of error, and by that we mean reduction of $\sigma_e^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$, is an important aspect of experimental design.

6.9.3 Comparison of CRDs without and with Subsampling

Another benefit of subsampling is that, even though it is not a substitute for replication, it may nevertheless lead to a reduction in the number of replications for the treatments, compared to a CRD without subsampling.

Table 6.11 ANOVA for CRD with Subsampling

Source	d.f.	SS	MS	$E(\text{MS})$
Treatments	$t - 1$	$r'n \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \text{SS}(T)$	$\text{MS}(T)$	$\sigma_\eta^2 + n\sigma_\varepsilon^2 + \frac{r'n}{t-1} \sum_i \tau_i^2$
Expt. Error	$t(r' - 1)$	$n \sum_{ij} (\bar{y}_{ij.} - \bar{y}_{i..})^2 = \text{SS}(EE)$	$\text{MS}(EE)$	$\sigma_\eta^2 + n\sigma_\varepsilon^2$
Obs. Error	$tr'(n - 1)$	$\sum_{ijk} (y_{ijk} - \bar{y}_{ij.})^2 = \text{SS}(OE)$	$\text{MS}(OE)$	σ_η^2
Total	$tr'n - 1$	$\sum_{ijk} (y_{ijk} - \bar{y}_{...})^2$		

We have seen in Section 6.8 that the number of replications r required may be quite large, in fact larger than is possible for practical and economical purposes. We may ask: How can subsampling be of some help? Suppose we can choose between two situations:

Plan I: CRD with r replications and no subsampling, that is, $r' = r$, $n = 1$.

Plan II: CRD with r' replications and subsampling of $n > 1$ OUs per EU with $r' < r$.

In plan I the F -test is based on $t(r - 1)$ d.f. in the denominator and the noncentrality parameter is

$$\lambda_I = \frac{r \sum \tau_i^2}{2(\sigma_\varepsilon^2 + \sigma_\eta^2)} \quad (6.50)$$

whereas for plan II the F -test is based on $t(r' - 1)$ d.f. in the denominator and the noncentrality parameter is

$$\lambda_{II} = \frac{r'n \sum \tau_i^2}{2(\sigma_\eta^2 + n\sigma_\varepsilon^2)} = \frac{r' \sum \tau_i^2}{2\left(\frac{\sigma_\eta^2}{n} + \sigma_\varepsilon^2\right)}. \quad (6.51)$$

Since the power of the F -test increases with the d.f. and the noncentrality parameter, plan II can be better than plan I only if $\lambda_{II} > \lambda_I$ since $t(r' - 1) < t(r - 1)$. Exactly what this relationship should be is hard to tell in general since this depends obviously on the values of r , r' , n , σ_ε^2 , σ_η^2 in a complex way. One way to look at this somewhat constructively is to compare $\text{var}(\bar{y}_{i..} - \bar{y}_{i'..})$, that is, the variance of a simple treatment comparison, for both situations. Specifically, this variance for plans I and II is given by

$$\text{var}_I = 2(\sigma_\varepsilon^2 + \sigma_\eta^2)/r \quad (6.52)$$

and

$$\begin{aligned} \text{var}_{II} &= 2(\sigma_\eta^2 + n\sigma_\varepsilon^2)/r'n \\ &= 2\left(\frac{\sigma_\eta^2}{r'n} + \frac{\sigma_\varepsilon^2}{r'}\right), \end{aligned} \quad (6.53)$$

respectively. One of the aims of experimental design is to reduce $\text{var}(\bar{y}_{i..} - \bar{y}_{i'..})$ as much as possible. Expression (6.53) shows clearly that this cannot be done by increasing n alone; that reduces only one component and usually the less important one at that. We, therefore, have to consider both r' and n carefully in our choice of the design. A useful relationship between r , r' , n can be obtained by equating (6.52) and (6.53) and letting $\sigma_\eta^2 = \delta\sigma_\varepsilon^2$. We find then that

$$n = \frac{r\delta}{r'(1 + \delta) - r} \quad (6.54)$$

or

$$r' = \frac{r(\delta + n)}{n(\delta + 1)}. \quad (6.55)$$

The way we may use these relationships is as follows:

- (i) Based on an appropriate choice of Δ^* [see (6.45)], find r from the Bowman and Kastenbaum (1975) tables (see Table 6.8). We note that Δ^* does not depend on the choice of the design, that is, the CRD without or with subsampling; in either case $\sigma_e = (\sigma_\epsilon^2 + \sigma_\eta^2)^{1/2}$;
- (ii) choose an r' in the neighborhood of r with $r' < r$;
- (iii) specify a value for δ based on empirical or theoretical evidence;
- (iv) use (6.54) to determine an appropriate n , rounding up to integer values.

We illustrate this procedure with the following example.

EXAMPLE 6.7: Suppose $t = 5$, $\alpha = .05$, $1 - \beta = .80$, $\Delta^* = 1.50$. From Table 6.8 we find $r = 12$. For $\delta = .50, .75, 1.00$ the possible choices of r' and n are given in Table 6.12. \square

The results of Table 6.12 show, that in general,

- (i) we only have a limited number of choices for r' ;
- (ii) as r' decreases, n increases rapidly;
- (iii) as δ increases, more options for r' are available;
- (iv) the total number of observations, $tr'n$, for the CRD with subsampling is always considerably larger than the total numbers of observations, tr , for the CRD without subsampling.

The important point of this whole discussion is that we must carefully evaluate our options before embarking on an experiment, taking the investigator's aims, the availability of experimental material, and limitation of resources into account. Only then can we avoid major disasters at the end of the experiment.

6.10 TRANSFORMATIONS

An important aspect of the analysis of experimental data is that of the scale of measurement. Problems can arise because of reasons mentioned below, namely nonadditivity of unit and treatment effects and nonconstancy of variances. Both phenomena are possibly related and may be resolved by transformation of the data to a more appropriate scale.

6.10.1 Nonadditivity in the General Sense

The reader will have noted that throughout we have made critical use of the idea of additivity of treatment contribution and unit contribution [see (6.3) and (6.16)]. This

Table 6.12 Numbers of Replications and Size of Subsamples for
 $r = 12$

$\delta = .50$		$\delta = .75$		$\delta = 1.00$	
r'	n	r'	n	r'	n
11	2	11	2	11	2
10	2	10	2	10	2
9	4	9	3	9	2
		8	5	8	3
		7	36	7	6

in essence amounts to a choice of scale of measurement. For example, if we have unit-treatment additivity for $\sqrt{T_{ik}}$, that is, $\sqrt{T_{ik}} = T_i + U_k$, then we obviously do not have additivity for

$$T_{ik} = T_i^2 + 2T_iU_k + U_k^2.$$

There has been extreme negligence with regard to this aspect of experimental inference. The problem was addressed by Neyman et al. (1935) and McCarthy (1937), but never by Fisher, which led to angry disagreements between Neyman and Fisher in the 1930s. Kempthorne (1952, Chapter 8) addressed this problem, but mainly in the context of the randomized complete block design. It is intrinsic in analyses of experimental data that additivity holds; otherwise different experiments will lead to the existence of interaction between experiments and treatments. It is, of course, not possible to establish for which scale of measurement additivity holds, but it is plausible to associate nonadditivity with nonconstancy of variances, and that can often be removed by a suitable transformation of the observations.

6.10.2 Nonconstancy of Variances

An essential aspect of the analysis of data from a CRD using model (6.19) or (6.46) is the constancy of variance of the observations. In general, this is not an unreasonable assumption. There exist, however, situations where this assumption is clearly not true. For example, in an experiment to compare different nutrients with respect to germination rate of certain seeds we have n seeds in each pot (EU) and for each seed the observation is $y = 1$ if the seed germinates and $y = 0$ if it does not germinate. If for the i th treatment the probability (rate) for germination is P_i , then obviously

$$\begin{aligned} E(y_{ijk}) &= P_i \\ \text{var}(y_{ijk}) &= P_i(1 - P_i) \end{aligned}$$

and

$$E(\bar{y}_{ij.}) = P_i$$

$$\text{var}(\bar{y}_{ij.}) = \frac{P_i(1 - P_i)}{n}.$$

Hence if there are differences among the treatments, then we do not have constancy of variances. This is an example where the variance is a function of the mean of the observations. More generally we can express this as follows. If y is the observation with $E(y) = \mu^*$, we then write

$$\text{var}(y) = g^2(\mu^*), \quad (6.56)$$

where $g(\mu^*)$ is some function of μ^* (we note here that in our case $\mu^* = \mu + \tau_i$ if the observation y is obtained for the i th treatment and that for this reason the $g(\mu^*)$ in (6.56) are possibly different). To equalize or stabilize the variance across treatments we use a transformation of the observations y , say $f(y)$, such that

$$\text{var}[f(y)] = \text{constant} = c^2, \text{ say.} \quad (6.57)$$

6.10.3 Choice of Transformation

To determine a suitable transformation we use the Taylor series expansion of $f(y)$ around μ^* (in the statistical literature this is also referred to as the *method of statistical differentials* or the *delta method*) and write

$$f(y) \cong f(\mu^*) + f'(\mu^*)(y - \mu^*) + \text{remainder} \quad (6.58)$$

Taking the expected value of both sides of (6.58) gives

$$E[f(y)] \cong f(\mu^*)$$

and hence, using (6.56) and (6.58),

$$\text{var}[f(y)] \cong [f'(\mu^*)]^2 g^2(\mu^*), \quad (6.59)$$

It follows then from (6.57) and (6.59) that

$$f'(\mu^*) = \frac{c}{g(\mu^*)} \quad \text{or} \quad f(\mu^*) = \int \frac{c}{g(\mu^*)} d\mu^*, \quad (6.60)$$

For observations for which (6.56) holds (at least approximately) because of theoretical considerations or because of empirical evidence (by plotting the residuals, for example), we can then determine an appropriate transformation $f(y)$ from (6.60). Examples of some well-known transformations (see Bartlett, 1947; Kempthorne, 1952, Section 8.5) are given in Table 6.13.

Table 6.13 Some Useful Transformations

No.	$g^2(\mu^*)$	y	$f(y)$	$\text{var}[f(y)]$	Parent Distribution
1	$\left. \begin{matrix} \mu^* \\ \lambda^2 \mu^* \end{matrix} \right\}$	Counts	\sqrt{y}	$\begin{cases} .25 \\ .25\lambda^2 \end{cases}$	Poisson Empirical
2	$\lambda^2 \mu^{*2}$	y_{ij}	$\log y$ $(\log(y+1))$	λ^2	Empirical
3	$\frac{2\mu^{*2}}{n-1}$	s_{ij}^2	$\log y$	$\frac{2}{n-1}$	Sample Variance
4	$\frac{\mu^*(1-\mu^*)}{n}$	Number of "Successes"	$\arcsin \sqrt{\frac{y}{n}}$	$\begin{cases} 821/n(\text{degrees}) \\ .25/n(\text{radians}) \end{cases}$	Binomial
5	$\lambda \mu^*(1-\mu^*)$	y_{ij}	$\arcsin \sqrt{y}$	$.25\lambda$ (radians)	Empirical
6	$\lambda^2 \mu^{*2}(1-\mu^*)^2$	y_{ij}	$\log \frac{y}{1-y}$	λ^2	Empirical
7	$\frac{(1-\mu^{*2})^2}{n-1}$	r_{ij}	$\frac{1}{2} \log \frac{1+y}{1-y}$	$\frac{1}{n-3}$	Sample Correlation
8	$\mu^* + \lambda^2 \mu^{*2}$	Counts	$\frac{1}{\lambda} \arcsin(\lambda \sqrt{y})$.25	Empirical

It is of course, obvious from (6.58) and (6.59) that the transformations given in Table 6.13 will not achieve complete constancy of variance. They serve, however, as basic transformations and are quite satisfactory from a practical point of view even if the form of $g(\mu^*)$ is not entirely correct. Modifications to some of the transformations in Table 6.13 have been proposed to enhance their performance, that is, come closer to achieving constancy. Freeman and Tukey (1950) proposed to replace transformation 1 in Table 6.13 for $n\mu^* \geq 1$ by

$$f(y) = \sqrt{y} + \sqrt{y+1}$$

with $\text{var}[f(y)] \cong 1$, and transformation 4 for $\mu^* > 1$ by

$$f(y) = \frac{1}{2} \left\{ \arcsin \sqrt{\frac{y}{n+1}} + \arcsin \sqrt{\frac{y+1}{n+1}} \right\}.$$

6.10.4 Power Transformations

For situations where there does not necessarily exist a relationship between the mean and the variance as discussed in the previous section, Box and Cox (1964) have proposed a parametric family of transformations:

$$f(y) \equiv \begin{cases} y(\lambda) = (y^\lambda - 1)/\lambda & (\lambda \neq 0) \\ y(0) = \log y \end{cases} \quad (6.61)$$

Because of the form of the transformations (6.61) they are also referred to as *power transformations*. The general idea here is to estimate λ from the data and then use $y(\hat{\lambda})$ as the actual transformation, where $\hat{\lambda}$ is the estimate of λ . Since the scale of the transformed observation depends on λ , that is, $\hat{\lambda}$, Box and Cox (1964) suggested to use the normalized transformation

$$\begin{aligned} z(\lambda) &= (y^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}) (\lambda \neq 0) \\ z(0) &= \dot{y} \log \dot{y} \end{aligned}$$

instead, where \dot{y} is the geometric mean of the observations y .

Since the primary objective of the transformations (6.61) is to achieve normality the estimator for λ is obtained by assuming a multivariate normal distribution for $y(\lambda)$ with constant variance, the secondary objective, for a “simple” linear model, the tertiary objective. With these assumptions, that is, objectives, $\hat{\lambda}$ can then be obtained using the theory of maximum likelihood. Also, approximate confidence limits for λ can be obtained giving the user a wider choice for $\hat{\lambda}$ which may be helpful in interpreting the transformation, that is, rather than using, for example, $\hat{\lambda} = -.9$ a more suitable, and perhaps plausible, choice may be to use $\hat{\lambda} = -1$. As normality is not a crucial assumption in our discussions, we shall not pursue this topic further but refer the reader to the Box and Cox (1964) results.

6.11 EXAMPLES USING SAS®

The aim of this chapter is to explain and describe the nature, philosophy, and properties of the CRD together with the underlying linear model and the associated analysis. The reader will have noticed that we did not pay any attention to purely computational issues. The reason for this is, of course, that statistical packages are available to perform the analysis conveniently and without any difficulty. We shall not mention the various computer packages available (their number rises almost daily) as every user has his or her favorite. Instead we shall present some examples and make some comments about the analysis using the Statistical Analysis System (SAS) (SAS Institute, Inc., 2002–2003) as a representative of many suitable programs. It is assumed that the reader has some familiarity with the fundamentals of SAS.

EXAMPLE 6.8: Consider an experiment to compare the efficiency of five different types of gasoline, measured in miles/gallon. Ten cars, all of the same make and model, are available. Each car is randomly assigned a particular type of gasoline such that each type is assigned to two cars. Each car is driven a specified route and the gas consumption, converted to miles/gallon, is recorded.

This is a CRD with $t = 5$ treatments and $r = 2$ replications. Each car is the EU and OU.

The analysis is performed using SAS PROC GLM. The input statements are given in Table 6.14a and the output is given in Table 6.14b.

We comment briefly on some aspects of the output:

- (i) It is useful to have the data set printed out, since it provides an easy and convenient way to detect obvious typographical errors that may have occurred in the data recording and/or input (we note here that in the following we may not always adhere to this suggestion in order to conserve space).
- (ii) The ANOVA table is of the form given in Table 6.2. For the CRD the Model SS and the Brand Type I and Type III SS are identical.
- (iii) The P -value of 0.0049 indicates that there are differences among the brands of gasoline.
- (iv) The brand least squares means are identical to the means. The P -values given here are of no interest as they test $H_{0i} : \mu + \tau_i = 0 (i = 1, 2, \dots, 5)$.

□

EXAMPLE 6.9: As an extension of the experiment described in Example 6.8 we consider now the situation where each car is driven three times and a measurement is obtained after each drive. We thus have a CRD with subsampling, the drives representing the subsamples. More precisely we then have $t = 5$, $r' = 2$, $n = 3$. The data are provided in Table 6.15a together with the SAS input statements using, mainly for purposes of comparison, SAS PROC GLM and SAS PROC MIXED. The output for both procedures is given in Table 6.15b.

Table 6.14 Data Analysis for CRD

a.) Input Statements:

```
data crdgas;
input brand miles @@;
datalines;
1 25.8 1 23.9
2 28.5 2 27.9
3 22.3 3 24.0
4 29.5 4 28.7
5 26.0 5 25.8
;
run;

proc print data=crdgas;
title1 'DATA FOR CRD (t=5, r=2)';
run;

proc glm data=crdgas;
class brand;
model miles=brand;
lsmeans brand/stderr;
title1 'ANALYSIS OF CRD';
run;
```

b.) Output:

DATA FOR CRD (t=5, r=2)

Obs	brand	miles
1	1	25.8
2	1	23.9
3	2	28.5
4	2	27.9
5	3	22.3
6	3	24.0
7	4	29.5
8	4	28.7
9	5	26.0
10	5	25.8

Table 6.14 (Continued)

ANALYSIS OF CRD						
The GLM Procedure						
Class Level Information						
Class	Levels	Values				
brand	5	1	2	3	4	5
Number of Observations Read		10				
Number of Observations Used						
Dependent Variable: miles						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	4	47.23400000	11.80850000	15.66	0.0049	
Error	5	3.77000000	0.75400000			
Corrected Total	9	51.00400000				
R-Square		Coeff Var	Root MSE	miles Mean		
0.926084		3.309191	0.868332	26.24000		
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
brand	4	47.23400000	11.80850000	15.66	0.0049	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
brand	4	47.23400000	11.80850000	15.66	0.0049	
Least Squares Means						
brand	miles LSMEAN	Standard Error	Pr > t			
1	24.8500000	0.6140033	<.0001			
2	28.2000000	0.6140033	<.0001			
3	23.1500000	0.6140033	<.0001			
4	29.1000000	0.6140033	<.0001			
5	25.9000000	0.6140033	<.0001			

We make the following comments concerning the output:
For PROC GLM:

- (i) The model SS contains both SS(brand) and SS(car within brand), the latter representing - in a technical sense - the experimental error;
- (ii) both sums of squares in (i) are separated in the Type I and Type III ANOVAs;
- (iii) the F -value and the P -value for brand under Type III ANOVA are incorrect and should be disregarded (see (iv) below);
- (iv) the correct F - and P -values for brand are provided after specifying in the input statement that, according to Table 6.11, $F = MS(\text{brand})/MS(\text{car (brand)}) = 33.13/1.36 = 24.48$ with $P = .0018$;
- (v) a statement has to be provided in the input to obtain the correct standard error for the least squares means such that, just as in (iv) above, $MS(\text{car(brand)})$ should be used as the appropriate $MS(\text{experimental error})$.

For PROC MIXED:

- (vi) In the input statement the source car(brand) is recognized as the experimental error by declaring the effect as a random effect;
- (vii) as a consequence of the statement in (vi) above estimates for σ_ϵ^2 and σ_η^2 will be obtained. The estimation method used is, by default, REML (REsidual Maximum Likelihood) (for details about REML see Section II.1.11.2). We obtain $\hat{\sigma}_\epsilon^2 = .39$ and $\hat{\sigma}_\eta^2 = .19$. For balanced data the REML estimates are the same as those given in (6.49) and (6.48), respectively;
- (viii) the test for no differences among brands is now performed correctly without further input. The same holds for the standard errors of the least squares means;
- (ix) the above comments show that PROC MIXED appears to be the preferred method. Only if we like to obtain the ANOVA table as given in Table 6.11 should we use PROC GLM. \square

6.12 EXERCISES

- 6.1** Consider the following results from a CRD with $t = 2$ treatments and $r = 4$ replications for each treatment:

EU #	1	2	3	4	5	6	7	8
Response	7	2	1	6	4	8	10	4
Treatment	2	1	1	1	2	2	2	1

- (i) Using the ratio $MS(T)/MS(E)$ as the test criterion perform the randomization test for $H_0 : \tau_1 = \tau_2$.

Table 6.15 Data Analysis for CRD with Subsampling

a.) Input Statements:

```
data mileage;
input brand car miles @@;
datalines;
1 1 25.8 1 1 25.6 1 1 26.0 1 2 23.9 1 2 24.2 1 2 23.5
2 1 28.5 2 1 28.0 2 1 28.4 2 2 27.9 2 2 28.1 2 2 28.4
3 1 22.3 3 1 22.7 3 1 23.0 3 2 24.0 3 2 23.1 3 2 23.5
4 1 29.5 4 1 27.5 4 1 29.1 4 2 28.7 4 2 29.0 4 2 28.8
5 1 26.0 5 1 25.7 5 1 26.1 5 2 25.8 5 2 25.6 5 2 26.3
;
run;

proc glm data=mileage;
class brand car;
model miles=brand car (brand);
test H=brand E=car (brand);
lsmeans brand/stderr E=car (brand);

title1 'ANALYSIS OF CRD W/SUBSAMPLING';
title2 'USING PROC GLM';
run;

proc mixed data=mileage;
class brand car;
model miles=brand;
random car (brand);
lsmeans brand;
title2 'USING PROC MIXED';
run;
```

b.) Output:

ANALYSIS OF CRD W/SUBSAMPLING						
USING PROC GLM						
The GLM Procedure						
Class Level Information						
Class	Levels		Values			
brand	5		1	2	3	4 5
car	2		1	2		
Number of Observations Read						30
Number of Observations Used						30

Table 6.15 (Continued)

Dependent Variable: miles

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	140.0466667	15.5607407	80.21	<.0001
Error	20	3.8800000	0.1940000		
Corrected Total		29	143.9266667		

R-Square	Coeff Var	Root MSE	miles Mean
0.973042	1.683265	0.440454	26.16667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
brand	4	133.2433333	33.3108333	171.71	<.0001
car(brand)	5	6.8033333	1.3606667	7.01	0.0006

Source	DF	Type III SS	Mean Square	F Value	Pr > F
brand	4	133.2433333	33.3108333	171.71	<.0001
car(brand)	5	6.8033333	1.3606667	7.01	0.0006

Tests of Hypotheses Using the Type III MS for car(brand) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
brand	4	133.2433333	33.3108333	24.48	0.0018

Least Squares Means

Standard Errors and Probabilities Calculated Using the Type III MS for car(brand) as an Error Term

brand	miles LSMEAN	Standard Error	Pr > t
1	24.8333333	0.4762119	<.0001
2	28.2166667	0.4762119	<.0001
3	23.1000000	0.4762119	<.0001
4	28.7666667	0.4762119	<.0001
5	25.9166667	0.4762119	<.0001

Table 6.15 (Continued)

ANALYSIS OF CRD W/SUBSAMPLING USING PROC MIXED				
The Mixed Procedure				
Model Information				
Data Set	WORK.MILEAGE			
Dependent Variable	miles			
Covariance Structure	Variance Components			
Estimation Method	REML			
Residual Variance Method	Profile			
Fixed Effects SE Method	Model-Based			
Degrees of Freedom Method	Containment			
Class Level Information				
Class	Levels	Values		
brand	5	1	2	3 4 5
car	2	1	2	
Iteration History				
Iteration	Evaluations	-2 Res Log Like		Criterion
0	1	58.65095075		0.00000000
1	1	48.64765549		
Convergence criteria met.				
Covariance Parameter Estimates				
Cov Parm		Estimate		
car(brand)		0.3889		
Residual		0.1940		
Fit Statistics				
-2 Res Log Likelihood		48.6		
AIC (smaller is better)		52.6		
AICC (smaller is better)		53.2		
BIC (smaller is better)		53.3		

Table 6.15 (Continued)

Type 3 Tests of Fixed Effects						
Effect		Num DF	Den DF	F Value	Pr > F	
brand		4	5	24.48	0.0018	
Least Squares Means						
Effect	brand	Estimate	Standard Error	DF	t Value	Pr > t
brand	1	24.8333	0.4762	5	52.15	<.0001
brand	2	28.2167	0.4762	5	59.25	<.0001
brand	3	23.1000	0.4762	5	48.51	<.0001
brand	4	28.7667	0.4762	5	60.41	<.0001
brand	5	25.9167	0.4762	5	54.42	<.0001

(ii) Compare the significance level achieved in (i) to that of the usual F -test.

6.2 A researcher has done a preliminary study in the form of a CRD with subsampling to help him decide on the final design. He wants to compare five (5) treatments. In the preliminary study he has used two (2) experimental units for each treatment and two (2) observations per experimental unit.

The partial ANOVA for the data from the preliminary study is given below:

Source	SS
Treatments	43.58
Expt. Error	55.00
Sampling Error	30.00

and the 5 treatment means are 10.00, 12.30, 11.80, 14.25, 13.56.

In the final experiment the researcher wants to compare the same 5 treatments. He wants to use a CRD with or without subsampling.

(i) Suppose he wants to detect, with probability .9, approximately the same difference between the best and worst treatment as observed in the study, using a test of size .05. Based on the information available from the preliminary study, how many replications does he need for a CRD without subsampling?

- (ii) Suppose the testing of hypotheses is not so important. He would like to consider possible CRDs with subsampling that achieve a variance of treatment comparisons no larger than the variance for the CRD obtained in (i). Give possible options of designs.
- 6.3** An agronomist conducted a field trial to compare the relative effects of five particular fertilizers on the yield of Trebi barley. Thirty homogeneous experimental plots are available and six were assigned at random to each fertilizer treatment. At harvest time, three sample quadrats were taken (at random) from each experimental plot and the yield was obtained for each of the 90 quadrats.
- (i) What is the name of the experimental design used? Give an appropriate model for analyzing the data from this experiment.
 - (ii) The agronomist has consulted two “statisticians” for the analysis of the data. He wants to know whether there are differences among the fertilizer effects. He is confused by the three SAS printouts (see Tables 6.16 a, b, c) and he comes to you for help to find out which analysis is correct. Based on the information provided explain how an appropriate test should be carried out.
 - (iii) What is the variance of a single observation [express as a formula based on your model statement in (i)] and how much of this variance is due to experimental error and to observational (sampling) error (use the numerical information provided).
 - (iv) Give the standard error for a simple treatment comparison.
- 6.4** A pharmaceutical company conducts an experiment to compare 5 drugs. 30 animals are available for the trial. Each drug is injected into 6 randomly selected animals. All the animals are very similar. After an appropriate period of time 2 blood samples are taken from each animal and duplicate analyses are made for each blood sample. The reading from each analysis represents the observation to be used for the statistical analysis of this experiment.
- (i) What kind of experimental design has been used?
 - (ii) What feature does this design have which we have not encountered before?
 - (iii) How many types of errors can you identify for this design? Give their names.
 - (iv) Write out an appropriate model for this design.
 - (v) Based upon this model, outline the ANOVA table, giving sources of variation and degrees of freedom.
 - (vi) Using the ANOVA table, indicate how you would test the hypothesis that there are no differences among the drugs.

6.5 For the CRD with t treatments, r' replications per treatment and n observations per EU, show that

$$\text{var} \left(\sum_i c_i \bar{y}_{i..} \right) = \sum_i c_i^2 \frac{\sigma_\eta^2 + n\sigma_\varepsilon^2}{r'n}.$$

Table 6.16 SAS Inputs and Outputs for Exercise 6.3

a.) SAS Code:

```
proc glm;
class ~fert;
model y = fert;
run;
```

SAS Printout (3A):

General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
FERT	5	1	2	3	4 5
Number of observations in data set = 90					
General Linear Models Procedure					
Dependent Variable: ~Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	65240.711	16310.178	242.72	0.0001
Error	85	5711.778	67.197		
Corrected Total	89	70952.489			
R-Square C.V. Root MSE Y Mean					
0.919499 10.20988 8.1974 80.289					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
FERT	4	65240.711	16310.178	242.72	0.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
FERT	4	65240.711	16310.178	242.72	0.0001

Table 6.16 (Continued)

b.) SAS Code:

```
proc glm;
  classes fert rep;
  model y = fert rep(fert);
run;
```

SAS Printout:

General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
FERT	5	1	2	3	4 5
REP	6	1	2	3	4 5 6
Number of observations in data set = 90					
General Linear Models Procedure					
Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	67139.822	2315.166	36.43	0.0001
Error	60	3812.667	63.544		
Corrected Total	89	70952.489			
R-Square		C.V.	Root MSE	Y Mean	
0.946265		9.928493	7.9715	80.289	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
FERT	4	65240.711	16310.178	256.67	0.0001
REP (FERT)	25	1899.111	75.964	1.20	0.2811
Source	DF	Type III SS	Mean Square	F Value	Pr > F
FERT	4	65240.711	16310.178	256.67	0.0001
REP (FERT)	25	1899.111	75.964	1.20	0.2811

c.) SAS Code:

```
proc glm;
  classes fert rep;
  model y = fert rep(fert);
  test h=fert e=rep(fert);
run;
```


Table 6.16 (Continued)

SAS Printout:

General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
FERT	5	1	2	3	4 5
REP	6	1	2	3	4 5 6

Number of observations in data set = 90

General Linear Models Procedure					
Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	67139.822	2315.166	36.43	0.0001
Error	60	3812.667	63.544		
Corrected Total	89	70952.489			
R-Square C.V. Root MSE Y Mean					
0.946265 9.928493 7.9715 80.289					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
FERT	4	65240.711	16310.178	256.67	0.0001
REP (FERT)	25	1899.111	75.964	1.20	0.2811
Source	DF	Type III SS	Mean Square	F Value	Pr > F
FERT	4	65240.711	16310.178	256.67	0.0001
REP (FERT)	25	1899.111	75.964	1.20	0.2811

Tests of Hypotheses using the Type III MS for REP (FERT) as an error term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FERT	4	65240.711	16310.178	214.71	0.0001

CHAPTER 7

Comparisons of Treatments

7.1 INTRODUCTION

We have mentioned earlier that the aim of any experiment is to compare treatments. One way to do this is in the context of the ANOVA by testing the hypothesis $H_0: \tau_1 = \tau_2 = \cdots = \tau_t = 0$. In most situations, however, the result from such a test is not very informative. To arrive at the conclusion that there are differences among the treatments (typically associated with a small P -value) is generally no major surprise. The immediate question then is: which treatments are different from each other, or how large is the difference between certain treatment effects, or is there a systematic pattern describing the magnitudes of treatment effects? As so often, the type of question(s) depends on the particular experiment and the kind of treatments used in the experiment.

It is obviously impossible to anticipate all conceivable questions. There are, however, certain kinds of comparisons that can be grouped into a number of categories based on the statistical methodology used to analyze such comparisons. We shall give a brief discussion of these methods in the context of the CRD, but the reader should have no difficulty adapting these methods to other designs, as described in later chapters, as well.

7.2 PREPLANNED COMPARISONS FOR QUALITATIVE TREATMENTS

In many experiments the t treatments under investigation bear some relationship to each other which immediately suggests that certain comparisons are of more interest than others. These comparisons are indeed often the basis for the experiment. They are referred to as *a priori comparisons* or *preplanned comparisons*.

7.2.1 Treatment Contrasts

As an example consider the following experiment. We want to compare the effectiveness of two different pesticides, A and B say, both applied in two different forms: spray (A_1 and B_1) and powder (A_2 and B_2). A control (that is, no pesticide), C say, is included in the experiment in order to establish any effectiveness of the pesticides at all. Thus we have altogether $t = 5$ treatments, C, A_1, A_2, B_1, B_2 , and each treatment is applied randomly to r more or less uniformly infested plots of land. The aim of the experiment and the “structure” of the treatments suggest the following comparisons:

- (i) control vs. pesticides,
- (ii) pesticide A vs. pesticide B ,
- (iii) application A_1 vs. A_2 ,
- (iv) application B_1 vs. B_2 ,
- (v) spray vs. powder,
- (vi) A_1 vs. B_1 ,
- (vii) A_2 vs. B_2 .

Using model (6.22) for the observations from this experiment, we can express any of the comparisons above in terms of the treatment effects as

$$C_l = \sum_{i=1}^5 c_{li} \tau_i \quad (7.1)$$

($l = 1, 2, \dots, 7$), where the c_{li} are constants such that $\sum_i c_{li} = 0$ for every l and the τ_i represent the treatment effects as indicated below:

i	1	2	3	4	5
Treatment	C	A_1	A_2	B_1	B_2

7.2.2 Orthogonal Contrasts

The coefficients c_{li} in (7.1) for the contrasts (i)–(vii) above are given in Table 7.1. A closer look at the coefficients for C_1, C_2, C_3, C_4 reveals that

$$\sum_i c_{li} c_{l'i} = 0 \quad (7.2)$$

for $l, l' = 1, 2, 3, 4$ and $l \neq l'$. Any two contrasts, C_l and $C_{l'}$, for which (7.2) is satisfied are called *orthogonal contrasts*. In this example then C_1, C_2, C_3 , and C_4 are orthogonal contrasts and so are C_1, C_5, C_6 , and C_7 . Each of these two sets of contrasts are referred to as a *complete set of orthogonal contrasts* for the five treatments.

Table 7.1 Coefficients for Contrasts

Contrast	i					Divisor
	1	2	3	4	5	
C_1	4	-1	-1	-1	-1	$\sqrt{20}$
C_2	0	1	1	-1	-1	2
C_3	0	1	-1	0	0	$\sqrt{2}$
C_4	0	0	0	1	-1	$\sqrt{2}$
C_5	0	1	-1	1	-1	2
C_6	0	1	0	-1	0	$\sqrt{2}$
C_7	0	0	1	0	-1	$\sqrt{2}$

In general, for t treatments there exist many (actually infinitely many) complete sets of $t - 1$ orthogonal contrasts each, but only few (if any) are useful for interpreting the results of the experiment. These are usually suggested by the structure of the treatments as in the example described above or as determined by the factorial structure of the treatments (see Chapter 11 and Chapter II.7).

7.2.3 Partitioning the Treatment Sum of Squares

A special feature of orthogonal contrasts is that they can be incorporated easily in the ANOVA in the sense that the sums of squares for the individual contrasts in a complete set provide a partitioning of the treatment sum of squares, $SS(T)$ in Table 6.1, into $t - 1$ single d.f. sums of squares. To show this, consider the contrast C_l as given in (7.1). An estimator for C_l is obviously

$$\hat{C}_l = \sum_i c_{li} \bar{y}_i. \quad (7.3)$$

with

$$\text{var}(\hat{C}_l) = \sum_i c_{li}^2 \frac{\sigma_e^2}{r}.$$

The sum of squares associated with C_l is then given by

$$SS(C_l) = \frac{[\hat{C}_l]^2}{\text{var}(\hat{C}_l)/\sigma_e^2} = \frac{r \left[\sum_i c_{li} \bar{y}_i \right]^2}{\sum_i c_{li}^2}. \quad (7.4)$$

We assume, without loss of generality, that $\sum_i c_{li}^2 = 1$ for $l = 1, 2, \dots, t - 1$. We refer to such comparisons as *standardized contrasts*. These can be obtained by replacing the coefficient c_{li} in (7.3) by $c_{li}^* = c_{li} / \sqrt{\sum_i c_{li}^2}$, for every $i = 1, 2, \dots, t$. We

then have $\sum_i (c_{li}^*)^2 = 1$. For example, for the contrast coefficients in Table 7.1 the standardizing divisors are given in the last column of Table 7.1. For ease of notation we drop in the following the asterisk in c_{li}^* .

Then (7.4) reduces to

$$SS(C_l) = r \left[\sum_i c_{li} \bar{y}_i \right]^2$$

or, if we write

$$\mathbf{c}'_l = (c_{l1}, c_{l2}, \dots, c_{lt})$$

and

$$\bar{\mathbf{y}}' = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_t)$$

it follows that

$$SS(C_l) = r[\mathbf{c}'_l \bar{\mathbf{y}}]^2. \quad (7.5)$$

To show that

$$SS(T) = \sum_{l=1}^{t-1} SS(C_l) \quad (7.6)$$

we consider the following set of t orthogonal linear functions of $\bar{\mathbf{y}}$ expressed in matrix notation as

$$\mathbf{z} = \begin{pmatrix} \mathbf{c}'_0 \\ \mathbf{c}'_1 \\ \vdots \\ \mathbf{c}'_{t-1} \end{pmatrix} \bar{\mathbf{y}} = \mathbf{C}' \bar{\mathbf{y}} \text{ say} \quad (7.7)$$

that is,

$$z_l = \mathbf{c}'_l \bar{\mathbf{y}} \quad (l = 0, 1, \dots, t-1)$$

with

$$\mathbf{c}'_0 = \frac{1}{\sqrt{t}} \mathbf{J}'$$

and \mathbf{J} being a vector of t unity elements. It follows then that \mathbf{C}' is an orthogonal matrix, that is,

$$\mathbf{C}' \mathbf{C} = \mathbf{I}$$

and hence

$$\mathbf{C} = (\mathbf{C}')^{-1},$$

which implies that

$$\mathbf{C} \mathbf{C}' = \mathbf{I}. \quad (7.8)$$

We then have from (7.7) and (7.8)

$$\mathbf{z}' \mathbf{z} = \bar{\mathbf{y}}' \mathbf{C} \mathbf{C}' \bar{\mathbf{y}} = \bar{\mathbf{y}}' \bar{\mathbf{y}} = \sum_i \bar{y}_i^2 \quad (7.9)$$

Table 7.2 ANOVA for CRD with Orthogonal Contrasts

Source	d.f.	SS	$E(MS)$
Treatments	$t - 1$	$SS(T)$	
C_1	1	$SS(C_1)$	$\sigma_e^2 + r(\mathbf{c}'_1\boldsymbol{\tau})^2$
C_2	1	$SS(C_2)$	$\sigma_e^2 + r(\mathbf{c}'_2\boldsymbol{\tau})^2$
\vdots	\vdots	\vdots	\vdots
C_{t-1}	1	$SS(C_{t-1})$	$\sigma_e^2 + r(\mathbf{c}'_{t-1}\boldsymbol{\tau})^2$
Error	$t(r - 1)$	$SS(E)$	σ_e^2

On the other hand, it follows from (7.7) that

$$\begin{aligned}
 \mathbf{z}'\mathbf{z} &= \sum_{l=0}^{t-1} z_l^2 = \sum_{l=0}^{t-1} [\mathbf{c}'_l \bar{\mathbf{y}}]^2 \\
 &= \frac{1}{t} \left[\sum_i \bar{y}_{i.} \right]^2 + \sum_{l=1}^{t-1} [\mathbf{c}'_l \bar{\mathbf{y}}]^2 \\
 &= t\bar{y}_{..}^2 + \sum_{l=1}^{t-1} [\mathbf{c}'_l \bar{\mathbf{y}}]^2
 \end{aligned} \tag{7.10}$$

Equating the right-hand sides of (7.9) and (7.10) we obtain

$$\begin{aligned}
 \sum_{l=1}^{t-1} [\mathbf{c}'_l \bar{\mathbf{y}}]^2 &= \sum_i \bar{y}_{i.}^2 - t\bar{y}_{..}^2 \\
 &= \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2.
 \end{aligned} \tag{7.11}$$

Multiplying both sides of (7.11) by r and making use of (7.5) we thus obtain (7.6) and hence Table 7.2.

This result can be incorporated into the ANOVA table by amending Table 6.1 as given in Table 7.2. This shows that in order to test

$$H_0: \mathbf{c}'_l \boldsymbol{\tau} = 0$$

with $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_t)'$, we use, as an approximation to the randomization test, the F -test

$$F = \frac{SS(C_l)}{MS(E)} \sim F_{1, t(r-1)}. \tag{7.12}$$

Alternatively, an approximate $(1 - \alpha)$ 100% confidence interval for $\mathbf{c}'_l \boldsymbol{\tau}$ is given by

$$\mathbf{c}'_l \bar{\mathbf{y}} \pm t_{1-\alpha/2, t(r-1)} \left(\frac{MS(E)}{r} \right)^{1/2} \tag{7.13}$$

or, more generally,

$$\widehat{C}_l \pm t_{1-\alpha/2, t(r-1)} \left[\widehat{\text{var}} \left(\widehat{C}_l \right) \right]^{1/2}.$$

We conclude this section by emphasizing that, although the structure of the treatments often suggest a complete set of orthogonal contrasts, one should not insist on considering only orthogonal contrasts. Other contrasts, as long as they are preplanned, may be more informative for explaining the results from an experiment.

7.3 ORTHOGONALITY AND ORTHOGONAL COMPARISONS

Having discussed the concept of orthogonal comparisons for the equal replication CRD it seems appropriate and necessary to give some formal description of the basic ideas of and connection between orthogonality and orthogonal comparisons.

Orthogonality is a concept that is usually and easily applied to vectors. Let $\mathbf{z}' = (z_1, z_2, \dots, z_m)$ represent an m -vector. Linear forms in \mathbf{z} may be represented by $\mathbf{c}'_i \mathbf{z}$ where $\mathbf{c}'_i = (c_{i1}, c_{i2}, \dots, c_{im})$. Two linear forms, $\mathbf{c}'_1 \mathbf{z}$ and $\mathbf{c}'_2 \mathbf{z}$, are said to be orthogonal if $\mathbf{c}'_1 \mathbf{c}_2 = 0$, that is, if the vectors \mathbf{c}_1 and \mathbf{c}_2 are orthogonal vectors.

This notion of orthogonality is used when we refer to orthogonal treatment contrasts as discussed in the previous section, for example, two contrasts, $\mathbf{c}'_1 \boldsymbol{\tau}$ and $\mathbf{c}'_2 \boldsymbol{\tau}$ with $\mathbf{c}'_1 \mathbf{J} = \mathbf{c}'_2 \mathbf{J} = 0$, are orthogonal if $\mathbf{c}'_1 \mathbf{c}_2 = 0$. But these comparisons also bear on random variables, namely, when we consider the estimated comparisons. For this we need to consider comparisons of random variables.

Let $\mathbf{x}' = (x_1, x_2, \dots, x_m)$ be random variables with variance-covariance matrix \mathbf{V} . Let $\mathbf{c}'_1 \mathbf{x}$ and $\mathbf{c}'_2 \mathbf{x}$ be linear functions of \mathbf{x} , with $\mathbf{c}'_1 \mathbf{J} = \mathbf{c}'_2 \mathbf{J} = 0$. The two comparisons, $\mathbf{c}'_1 \mathbf{x}$ and $\mathbf{c}'_2 \mathbf{x}$, are said to be orthogonal if $\text{cov}(\mathbf{c}'_1 \mathbf{x}, \mathbf{c}'_2 \mathbf{x}) = 0$, which is equivalent to $\mathbf{c}'_1 \mathbf{V} \mathbf{c}_2 = 0$. The term *orthogonal* as used with random variables, z_1 and z_2 say, means that $\text{cov}(z_1, z_2) = 0$. It is unfortunate that the adjective orthogonal was taken over from the mathematics of inner product spaces to the lack of covariance of random variables (just as, indeed, the term independence was taken over and leads to confusion).

Returning to estimated treatment comparisons, using observations from a CRD, we need to consider functions of random variables of the form $\mathbf{c}'_1 \bar{\mathbf{y}}$ and $\mathbf{c}'_2 \bar{\mathbf{y}}$, where $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_t)'$. We have shown in Section 6.3 that the variance-covariance matrix for $\bar{\mathbf{y}}$, using (6.13) and (6.14), is given by

$$\mathbf{V} = \left(\frac{1}{r} \mathbf{I} - \frac{1}{N} \mathbf{J} \mathbf{J}' \right) \sigma_u^2$$

for additivity in the strict sense, and

$$\mathbf{V} = \frac{1}{r} \mathbf{I} (\sigma_u^2 + \sigma_v^2 + \sigma_\eta^2) - \frac{1}{N} \mathbf{J} \mathbf{J}' \sigma_u^2$$

for additivity in the broad sense. It follows then immediately for the equal-replication CRD that if two treatment contrasts, $\mathbf{c}'_1 \boldsymbol{\tau}$ and $\mathbf{c}'_2 \boldsymbol{\tau}$, are orthogonal then the estimated contrasts, $\mathbf{c}'_1 \bar{\mathbf{y}}$ and $\mathbf{c}'_2 \bar{\mathbf{y}}$, are also orthogonal, that is, satisfy $\mathbf{c}'_1 \mathbf{V} \mathbf{c}_2 = 0$.

This result does, however, not hold for the unequal-replication CRD as considered in Section 6.7. The important change in \mathbf{V} as given above is that $(1/r)\mathbf{I}$ has to be replaced by \mathbf{R}^{-1} , where

$$\mathbf{R}^{-1} = \begin{pmatrix} \frac{1}{r_1} & & & 0 \\ & \frac{1}{r_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{r_t} \end{pmatrix}.$$

It is then obvious that $\mathbf{c}'_1 \mathbf{V} \mathbf{c}_2 \neq 0$.

The implications of these results are that for the equal-replication CRD the contrast SSs are orthogonal (see Table 7.2), but for the unequal-replication CRD they are not, that is, (7.6) does not hold.

7.4 COMPARISONS FOR QUANTITATIVE TREATMENTS

As mentioned earlier, orthogonal contrasts are often suggested a priori by the treatment structure. This should not, however, be taken as a general rule. Other preplanned (and hence nonorthogonal) comparisons may be more suitable for answering the investigator's questions. The only difference to the discussion above is that (7.6) does not hold, but tests of the form (7.12) and confidence interval estimation of the form (7.13) can still be carried out. The main point here is that the comparisons are preplanned in a meaningful way, that is, determined by the intent of the experiment and not by the outcome of the experiment, and that the number of such comparisons is generally small.

7.4.1 Comparisons for Treatments with Equidistant Levels

Another type of preplanned comparison arises if the treatments represent quantitative levels of some factor, for instance, increasing amounts of fertilizer. Rather than compare individual levels with each other, it is more informative to investigate whether there exist certain trends in response to the treatments, for example, whether the increase (decrease) is linear or whether there exists some curvature.

Suppose x_1, x_2, \dots, x_t represent the t levels such that $x_1 = 0, x_2 = d, x_3 = 2d, \dots, x_t = (t-1)d$ with $d > 0$, that is, the levels are equidistant, with $\bar{x} = (t-1)d/2$. Without loss of generality we may take $d = 1$ and obtain

$$z_i = x_i - \frac{t-1}{2} \quad (7.14)$$

as the centered levels, with $\sum_i z_i = 0$. We then consider the contrast

$$C_L = \sum_{i=1}^t \frac{z_i}{\sqrt{\sum_{i'} z_{i'}^2}} \tau_i, \quad (7.15)$$

which is of the form (7.1) with $c_{Li} = z_i / \sqrt{\sum z_{i'}^2}$. The estimator for (7.15) is given by

$$\hat{C}_L = \sum_i \frac{z_i}{\sqrt{\sum_{i'} z_{i'}^2}} \bar{y}_i. \quad (7.16)$$

Using the fact that $\sum z_i = 0$ we recognize immediately that (7.16) is the estimate of the linear regression coefficient in the model

$$\bar{y}_i = \alpha + \beta w_i + \text{error}$$

with $w_i = z_i / \sqrt{\sum z_{i'}^2}$, that is, $\hat{\beta} = \hat{C}_L$. Alternatively, for the models

$$\bar{y}_i = \alpha + \beta^* z_i + \text{error} \quad (7.17)$$

or

$$\bar{y}_i = \alpha^* + \beta^* x_i + \text{error}$$

the estimator for β^* is $\hat{\beta}^* = \hat{C}_L / \sqrt{\sum z_i^2}$. Thus, in either case, \hat{C}_L is a measure of the linear increase (decrease) due to the increasing treatments.

7.4.2 Use of Orthogonal Polynomials

Model (7.17) may not describe adequately the relationship between treatments and responses. We may, for example, want to allow for curvature in the response by considering a model of the form

$$\bar{y}_i = \alpha + \beta_1 z_i + \beta_2 z_i^2 + \text{error}. \quad (7.18)$$

This can be done by using the methods of regression analysis, that is, estimate α, β_1, β_2 by the method of least squares (see Chapter 4). If the emphasis, however, is to find out whether curvature exists and of what kind it is, then it is more convenient to rewrite (7.18) in terms of so-called *orthogonal polynomials* as

$$\bar{y}_i = \gamma_0 P_0(z_i) + \gamma_1 P_1(z_i) + \gamma_2 P_2(z_i) + \text{error}$$

or, more generally,

$$\bar{y}_i = \sum_{l=0}^{t-1} \gamma_l P_l(z_i) + \text{error}. \quad (7.19)$$

Here the $P_l(z_i)$ are polynomials of degree l ($l = 0, 1, \dots, t-1$) such that

$$\sum_{i=1}^t P_l(z_i) = 0 \quad (l = 1, 2, \dots, t-1) \quad (7.20)$$

$$\sum_{i=1}^t P_l(z_i)P_{l'}(z_i) = 0 \quad (l \neq l'). \quad (7.21)$$

The $P_l(z_i)$ are, apart from a constant factor, Tchebycheff polynomials used in statistics first by Fisher (1921). Specifically, for $l = 0, 1, 2, 3, 4$ we have

$$P_0(z_i) = 1$$

$$P_1(z_i) = \lambda_1 z_i$$

$$P_2(z_i) = \lambda_2 \left[z_i^2 - \frac{1}{12}(t^2 - 1) \right]$$

$$P_3(z_i) = \lambda_3 \left[z_i^3 - \frac{1}{20}(t^2 - 7)z_i \right]$$

$$P_4(z_i) = \lambda_4 \left[z_i^4 - \frac{1}{14}(3t^2 - 13)z_i^2 + \frac{3}{560}(t^2 - 1)(t^2 - 9) \right],$$

where the λ_l are chosen such that the $P_l(z_i)$ are positive or negative integers. These polynomials have been tabulated by, for example, Fisher and Yates (1957), Pearson and Hartley (1970), and Beyer (1991) for $t = 3, 4, \dots$, and $l = 2, 3, 4, 5, 6$. For $t = 3, 4, 5, 6$, the $P_l(z_i)$ are given in Table 7.3 (for $l = 1, 2, 3, 4$ these values can be obtained by substituting the z_i of (7.4) into the above expressions).

The estimator for the regression coefficients γ_l in (7.19) are obtained by the method of least squares as

$$\hat{\gamma}_l = \frac{\sum_i \bar{y}_i P_l(z_i)}{\sum_i [P_l(z_i)]^2} \quad (l = 0, 1, 2, \dots, t-1) \quad (7.22)$$

with

$$\text{var}(\hat{\gamma}_l) = \frac{\sigma_e^2}{r \sum_i [P_l(z_i)]^2}. \quad (7.23)$$

We can deduce easily, using (7.22) for $l = 1$ together with $P_1(z_i) = \lambda_1 z_i$, that

$$\hat{\gamma}_1 = \frac{1}{\lambda_1} \hat{\beta}^* = \frac{\hat{C}_L}{\lambda_1 \sqrt{\Sigma z_i^2}}$$

Table 7.3 Orthogonal Polynomials

$t=3$			$t=4$			$t=5$				$t=6$				
i	P_1	P_2	P_1	P_2	P_3	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4	P_5
1	-1	1	-3	1	-1	-2	2	-1	1	-5	5	-5	1	-1
2	0	-2	-1	-1	3	-1	-1	2	-4	-3	-1	7	-3	5
3	1	1	1	-1	-3	0	-2	0	6	-1	-4	4	2	-10
4			3	1	1	1	-1	-2	-4	1	-4	-4	2	10
5						2	2	1	1	3	-1	-7	-3	-5
6										5	5	5	1	1
λ	1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$	2	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{21}{10}$

Since $\lambda_1 \sqrt{\sum z_i^2}$ is a constant it follows then, of course, that $\hat{\gamma}_1$, too, is a measure of the linear trend of the response variable. Just as γ_1 represents the linear trend or contrast, γ_2 represents the quadratic contrast, γ_3 the cubic contrast, and so on. One of the important properties of the formulation (7.19) in terms of orthogonal polynomials is that the estimates $\hat{\gamma}_l$ are independent of the number of polynomials included in (7.19), for instance, $\hat{\gamma}_1$ is the same whether (7.19) is a linear model or a quadratic model or a cubic model, and so on.

7.4.3 Contrast Sums of Squares and the ANOVA

Since the orthogonal polynomials represent the coefficients of a complete set of orthogonal contrasts, it follows further that, as described earlier, this can be used to partition $SS(T)$ into $t - 1$ single d.f. sums of squares associated with $\gamma_1, \gamma_2, \dots, \gamma_{t-1}$, respectively. The sum of squares associated with γ_l , $SS(\gamma_l)$ say, is

$$SS(\gamma_l) = r(\hat{\gamma}_l)^2 \sum_i [P_l(z_i)]^2 \quad (7.24)$$

($l = 1, 2, \dots, t - 1$) using the rule for a single d.f. sum of squares. Then

$$SS(T) = \sum_{l=1}^{t-1} SS(\gamma_l). \quad (7.25)$$

To investigate the trend of the treatment effects, we would typically fit a low order model first, then check for lack of fit (LOF), and, if necessary, add further terms. Suppose we fit first the model

$$\bar{y}_i = \sum_{l=0}^q \gamma_l P_l(z_i) + \text{error} \quad (7.26)$$

with $q < t - 1$, then the ANOVA takes the form as given in Table 7.4, with

$$SS(LOF) = SS(T) - \sum_{l=1}^q SS(\gamma_l).$$

Then

$$F = \frac{MS(LOF)}{MS(E)} \sim F_{t-q-1, t(r-1)} \quad (7.27)$$

can be used to test whether (7.26) provides an adequate fit to the data. If the F -value in (7.27) is significant at a given level, then there is lack of fit and additional terms should be added to the model repeating the procedure just described.

The method of fitting orthogonal polynomials is quite easy to carry out. It depends, however, heavily on the fact that the levels of the treatment factor are equidistant, because this is the reason why orthogonal polynomials can be tabulated. For nonequidistant levels the method can be used also, except that one has to compute the polynomials sequentially, using the properties (7.20) and (7.21) (for a computational method see, for

Table 7.4 ANOVA for CRD Fitting Orthogonal Polynomials

Source	d.f.	SS
Treatments	$t - 1$	$SS(T)$
γ_1	1	$SS(\gamma_1)$
γ_2	1	$SS(\gamma_2)$
\vdots	\vdots	\vdots
γ_q	1	$SS(\gamma_q)$
Lack of fit	$t - q - 1$	Difference = $SS(LOF)$
Error	$t(r - 1)$	$SS(E)$

example, Narula, 1978). This is rather cumbersome and worthwhile only if the same experiment is being used repeatedly. It is for this reason that equidistant levels (actual levels or transformed levels such as log dose) should be used wherever possible.

In concluding this section we should emphasize that fitting polynomials may not always be appropriate. One may, for example, be interested in finding an asymptote to the response curve and then a model of the form

$$\bar{y}_{i.} = \alpha + \beta e^{-\gamma z_i} + \text{error}$$

or some variation of it might be used. The important point is that the nature and aims of the experiment should provide guidance for the analysis (see Chapter 2).

7.5 MULTIPLE COMPARISON PROCEDURES

7.5.1 Multiple Comparisons and Error Rates

It is rare that in a well thought out experiment the treatments do not have any structure which would lead naturally to the types of comparisons discussed in Sections 7.2 and 7.4 or Chapter 11. Occasionally, however, it may be desirable to make a large number of comparisons, for example, all possible pairwise comparisons, or comparisons suggested by the data (data-snooping). Whether one talks in terms of hypothesis testing or interval estimation, great care must be taken to use correct inference procedures. The major problem here is the so-called *multiplicity effect* (Tukey, 1977) which may lead to too many significant tests if incorrect procedures are used. The problem revolves around the notions of *comparisonwise error rate* (CWE), *familywise error rate* (FWE), and *per family error rate* (PFE). Basically, the CWE is being used for situations discussed in Section 7.2, whereas the FWE or PFE are being controlled in the types of comparisons mentioned above. Both FWE and PFE take the number of comparisons to be made (these constitute the “family”), N say, into account. If we

have $t = 10$ treatments, for example, we have $N = \binom{10}{2} = 45$ pairwise comparisons. The idea then is to control the error rate for this family rather than for each individual comparison, where error here refers to the Type I error in the context of hypothesis testing. The relationship between CWE and FWE can be expressed as (Hochberg and Tamhane, 1987; Westfall et al., 1999)

$$1 - \text{FWE} = (1 - \text{CWE})^N$$

or

$$\text{FWE} = 1 - (1 - \text{CWE})^N. \quad (7.28)$$

In order to achieve a certain FWE (say .10), we can use (7.28) to determine the CWE to be used for each individual test. As an approximation for small CWE and not too large N , we have

$$\text{FWE} \cong N \times \text{CWE}. \quad (7.29)$$

A number of multiple comparison procedures (MCPs), sometimes also referred to as *post-hoc comparisons*, have been developed which control the FWE either in the weak sense (under $H_0: \tau_1 = \tau_2 = \dots = \tau_t$ only) or in the strong sense (that is, under all configurations of the τ_i 's) (Hochberg and Tamhane, 1987). Most of these procedures have been available for some time and are widely used, but their acceptance is not universal. We, too, caution against their uncritical use but at the same time we feel that MCPs can play a useful role in data analysis. In what follows we describe briefly a few MCPs, but defer details to books on this subject such as Miller (1981), Hochberg and Tamhane (1987), Hsu (1996), Westfall et al. (1999).

Although all MCPs test the same null hypothesis they use different methods for controlling the FWE. They also differ in their sensitivity to alternative hypotheses. For these reasons different MCPs applied to the same data may give different results. Studies have been undertaken to compare the different MCPs (see Carmer and Swanson, 1973; Miller (1981); Stoline, 1981) but sometimes seemingly contradictory results make it difficult to give general recommendations which MCP to use in a given situation.

7.5.2 Least Significant Difference Test

This test was first proposed by Fisher (1935) and is now often referred to as *Fisher's protected LSD test*. It has two stages. At stage I we test $H_0: \tau_1 = \tau_2 = \dots = \tau_t = 0$ by the F -test of size α . If the F -test is nonsignificant, we terminate the analysis. If the F -test is significant, then at stage II we test each single comparison $H_0: \tau_i = \tau_{i'}$ by an α -level t -test with $t(r - 1)$ d.f.

7.5.3 Bonferroni t -Statistics

This test is based on the Bonferroni inequality

$$1 - \text{FWE} \geq 1 - N \times \text{CWE}.$$

Making use of (7.29), it uses α^* -level t -tests with $t(r-1)$ d.f. for the individual tests $H_0: \tau_i = \tau_{i'}$ where

$$\alpha^* = \alpha/N$$

and α is the chosen FWE. Tables for the upper percentage point of the Bonferroni t -statistic, $t_{\alpha/2N, t(r-1)}$ are given by Miller (1981).

7.5.4 Studentized Range Procedure

This procedure was proposed by Tukey (1952, unpublished). To test the hypotheses $H_0: \tau_i = \tau_{i'} (i \neq i')$ we compare $|\bar{y}_{i.} - \bar{y}_{i'.}|$ with $Q_{\alpha, t, t(r-1)} \sqrt{\text{MS}(E)/r}$, where $Q_{\alpha, t, t(r-1)}$ is the upper α 100% point of the studentized range distribution for t independent, unit normal, random variables divided by the square root of an independent χ^2_ν/ν random variable with $\nu = t(r-1)$ d.f. If, for a given α ,

$$|\bar{y}_{i.} - \bar{y}_{i'.}| \geq Q_{\alpha, t, t(r-1)} \sqrt{\frac{\text{MS}(E)}{r}} \quad (7.30a)$$

then τ_i and $\tau_{i'}$ are considered to be different from each other. Tables for $Q_{\alpha, t, t(r-1)}$ are given by e.g., Harter (1960), Hochberg and Tamhane (1987).

The MCP described above has been extended by Kramer (1956) to accommodate unequal numbers of replications. This method is referred to as the *Tukey-Kramer method*. It simply replaces (7.30a) by

$$|\bar{y}_{i.} - \bar{y}_{i'.}| \geq Q_{\alpha, t, \nu} \sqrt{\frac{1}{2} \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \text{MS}(E)}, \quad (7.30b)$$

where $\nu = \Sigma r_i - t$.

The test procedures given by (7.30a) and (7.30b) control the FWE at a given α . It is then easy to see that simultaneous $(1 - \alpha)$ 100% confidence intervals for all comparisons $\tau_i - \tau_{i'}$ can be obtained as

$$\bar{y}_{i.} - \bar{y}_{i'.} \pm Q_{\alpha, t, \nu} \sqrt{\frac{1}{2} \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \text{MS}(E)}.$$

7.5.5 Duncan's Multiple Range Test

This test was developed by Duncan (1951, 1955) to specifically test all hypotheses $H_0: \tau_i = \tau_{i'}$ by considering different error rates depending on the range of the two corresponding treatment means $\bar{y}_{i.}$ and $\bar{y}_{i'.}$. If, among the ordered treatment means, $\bar{y}_{i.}$ and $\bar{y}_{i'.}$ are p means apart, then an α_p -level studentized range test is conducted comparing $|\bar{y}_{i.} - \bar{y}_{i'.}|/\sqrt{\text{MS}(E)/r}$ with the critical value $Q_{\alpha_p, t, t(r-1)}$, where

$$\alpha_p = 1 - (1 - \alpha)^{p-1}.$$

We start by arranging the t treatment means in ascending order, say

$$\bar{y}_{[1]}, \bar{y}_{[2]}, \dots, \bar{y}_{[t]}$$

and compare $\bar{y}_{[t]} - \bar{y}_{[1]}$ with $Q_{\alpha_t, t, t(r-1)} \sqrt{MS(E)/r}$. If this difference is nonsignificant then all other differences are also nonsignificant; if the difference is significant, however, then $(t-1)$ -differences $\bar{y}_{[t]} - \bar{y}_{[2]}$ and $\bar{y}_{[t-1]} - \bar{y}_{[1]}$ are considered and compared with $Q_{\alpha_{t-1}, t-1, t(r-1)} \sqrt{MS(E)/r}$. This procedure is continued, reducing the critical value at each step, until no more differences can be declared significant. Tables for the critical values are given by Harter (1960) (see also Miller, 1981).

7.5.6 Scheffé's Procedure

One of the most general MCPs is that proposed by Scheffé (1953) for judging all possible contrasts in the analysis of variance setting. It is especially useful for data-snooping after the F -test for $H_0: \tau_1 = \tau_2 = \dots = \tau_t$ has been found significant, because the FWE for all possible contrasts is equal to α , the size of the F -test.

To test the hypothesis

$$H_0: \sum_{i=1}^t c_i \tau_i = 0$$

for all (c_1, c_2, \dots, c_t) such that $\sum c_i = 0$, we compare

$$\sum_i c_i \bar{y}_i / \sqrt{\sum_i c_i^2 \frac{MS(E)}{r}}$$

with the critical value

$$[(t-1)F_{1-\alpha, t-1, t(r-1)}]^{1/2}.$$

Alternatively, this procedure can be used to construct simultaneous $(1-\alpha)$ 100% confidence intervals for all contrasts $\sum c_i \tau_i$ in the form

$$\sum_i c_i \bar{y}_i \pm [(t-1)F_{1-\alpha, t-1, t(r-1)}]^{1/2} \times \left[\sum_i c_i^2 \frac{MS(E)}{r} \right]^{1/2}.$$

7.5.7 Comparisons with a Control

In some experiments the t treatments may consist of a control treatment and $t-1$ what may be referred to as test treatments, and the aim of the experiment is to compare the test treatments against the control (see also Section 9.8.2 and II.6.5). If treatment 1 is the control, then testing the hypotheses

$$H_0: \tau_1 = \tau_i \quad (i = 2, 3, \dots, t)$$

with FWE = α can be achieved by a procedure due to Dunnett (1955, 1964). Rather than compare the usual t -statistic

$$(\bar{y}_1 - \bar{y}_i) / \sqrt{2 \frac{MS(E)}{r}} \tag{7.31}$$

against the critical value of the t -distribution, we compare (7.31) with the critical value $|D|_{\alpha, t-1, t(r-1), \rho}$ for two-sided tests and $D_{\alpha, t-1, t(r-1), \rho}$ for one-sided tests. Tables

of these critical values are given by Hochberg and Tamhane (1987, Tables 5 and 4, respectively) using $\rho = .5$ (here ρ is the correlation between $\bar{y}_{1.} - \bar{y}_{i.}$ and $\bar{y}_{1.} - \bar{y}_{i'}$).

So far we have discussed the CRD with the same number of replications for all treatments. The situation described above, that is, one control, $t - 1$ test treatments, is one where it may be advisable to have unequal numbers of replications for the treatments. Specifically, we may use r replications for the test treatments and $r_c (> r)$ replications for the control in order to estimate $\mu + \tau_1$ as precisely as possible. For this case, the test procedure remains the same, except that the critical values change for two reasons: (i) The d.f. for $MS(E)$ are now $\nu = r_c - 1 + (t - 1)(r - 1)$, and (ii) the correlation coefficient between $\bar{y}_{1.} - \bar{y}_{i.}$ and $\bar{y}_{1.} - \bar{y}_{i'}$ is now given by

$$\rho^* = \frac{r}{r_c + r} < .5.$$

Critical values $|D|_{\alpha, t-1, \nu, \rho^*}$ and $D_{\alpha, t-1, \nu, \rho^*}$ are given by Hochberg and Tamhane (1987) for $\rho^* = .1$ and $.3$. Dunnett (1964) also provides a method for approximating the critical values for values of ρ^* other than those given above and for values $\rho_{ii'}$ where treatment i is replicated r_i times and not all r_i equal.

It should be clear, of course, that Dunnett's procedure cannot only be used for testing but also for obtaining $(1 - \alpha)$ 100% simultaneous confidence intervals for $\tau_1 - \tau_i (i = 2, 3, \dots, t)$.

7.5.8 Alternatives to Tests Based on Normality

Throughout our discussion in Chapter 6 of the analysis of data from a CRD, we have not made any assumptions about the underlying distribution of the data. We have relied on the approximation of the F -test to the randomization test to carry out tests in the ANOVA and we have similarly used indications of a computational nature to use the t -test as an approximation to the corresponding randomization test for follow-up studies (such as discussed in Section 7.2). The implication, of course, is that the assumption of normality is not generally of crucial importance for such tests (see also Scheffé, 1959).

The procedures discussed in this section all depend on the assumption of normality, and there are strong indications that they are not very robust against deviations from normality (Ringland, 1983). This is especially true for the Bonferroni procedure and less so for the Scheffé procedure. One alternative in such situations is to use nonparametric procedures. We shall not discuss this here any further, but nonparametric analogs to some of the procedures described here are presented and discussed by Miller (1981) and Hochberg and Tamhane (1987). One of the problems with MCPs is that in many cases they lead to results that are not easily interpreted (see also Section 7.4) and this problem may become even worse using nonparametric MCPs. In both cases the choice of the error rate α will be important, say $\alpha = .10$ or even $\alpha = .20$, certainly larger than the conventional $\alpha = .05$.

Another alternative to nonnormal situations is to use robust estimators for the treatment effects, such as M -estimators (Huber, 1981). This, however, leads to great difficulties in that the distributions of the test statistics will be difficult, if not impossible, to obtain and one would have to rely on Monte Carlo simulations or asymptotic results. These prospects together with the general difficulties of MCPs are not very promising.

7.6 GROUPING TREATMENTS

One of the objectives of MCPs, apart from comparing every treatment with every other, is to arrive at groups of “homogeneous” treatments. This will facilitate the interpretation of the results from the experiment and help in making recommendations concerning further action. Unfortunately, the picture is not always clear. Often we find overlapping of groups of treatments which are judged (according to some MCP) to be not significantly different from each other.

To illustrate this phenomenon, consider the following example from Snedecor (1946). During cooking, doughnuts absorb fats in various amounts. An experiment was done to investigate whether the amount of fat absorbed is different for different fats used. Eight fats (treatments) were compared, each with six mixes (replications). The treatment (fat) means, that is, the average amount of fat (in grams) absorbed by 24 doughnuts, are as follows:

Fat #	1	2	3	4	5	6	7	8
6pt] height \bar{y}_i .	172	178	182	185	165	176	161	162

If we order the treatment means and perform the studentized range test [see Section 7.5.4] with $\alpha = .10$, we obtain the following result (with $MS(E) = 141.6$ and $Q_{.10,8,40} = 4.099$, the critical value is 19.91).

Fat #	7	8	5	1	6	2	3	4
\bar{y}_i .	161	162	165	172	176	178	182	185

Here, underlined treatment means are not significantly different from each other. Based on these tests we can say, for example, that fats # 7,8 are different from fats # 3,4 with 7 and 8, and 3 and 4 not being different from each other. This does not mean, however, that (7,8) and (3,4) form two distinct groups, because 7 and 8 are not different from 5, 1, 6, 2, and 3 and 4 are not different from 1, 6, 2; in fact 3 is not different from 5 either. In summary, if one wants to establish groups of similar fats (for dietary purposes, for example) then it is clear that using the multiple range test (or other MCPs) will not accomplish that. What is needed is a procedure that uses, in combination with hypothesis testing, ideas of cluster analysis.

Such methods were developed by Scott and Knott (1974) and by Calinski and Corsten (1985). We shall describe here one of the methods proposed by Calinski and Corsten (1985) which is based on an extension of the studentized range procedure (see Section 7.5.4).

This procedure is a stepwise procedure and is referred to as a hierarchical, agglomerative procedure which uses ordinary distance as a working criterion. The adjective “hierarchical” means that once a treatment mean is included in a homogeneous cluster it will not be deleted in a subsequent step, and the adjective “agglomerative” means

that at each step two adjacent clusters (each or both consisting possibly of one element only) are combined to form a new cluster.

The algorithm starts with t clusters, namely the t treatments, represented by the treatment means, \bar{y}_i , ($i = 1, 2, \dots, t$), arranged in increasing order. At the first step the two closest treatments, as measured by the smallest $|\bar{y}_i - \bar{y}_{i'}|$, are combined to form one cluster and the range $R_1 = |\bar{y}_i - \bar{y}_{i'}|$ is compared with the critical value $C_\alpha = Q_{\alpha, t, t(r-1)} \sqrt{\text{MS}(E)/r}$ for a given α . At each following step a new cluster is formed by combining two adjacent clusters with the smallest range. The range R_s at step s ($1 \leq s \leq t-1$) is then compared with the critical value C_α . If $R_s > C_\alpha$ then the process stops and the clustering obtained at step $s-1$ will be the accepted grouping of the treatments. The groups thus formed are considered to be internally homogeneous with the studentized range test at significance level α .

We illustrate this procedure in Figure 7.1 for the example given above. Choosing $\alpha = .10$, the critical value is

$$C_{.10} = 4.099 \cdot \sqrt{\frac{141.6}{6}} = 19.91$$

and from Figure 7.1 it can be seen that the process stops at step 7 since $R_7 = 24 > C_{.10} = 19.91$. Hence the grouping arrived at prior to step 7 will be accepted, that is, fats 5, 7, 8 form one group and fats 1, 2, 3, 4, 6 form another group.

By using a fixed α , the probability of terminating too early, and hence accepting too many homogeneous groups is bounded by α . For small α this may lead, in fact, to too few groups. Here, as with MCPs in general, the choice of α is important and an α of .10 or .20 may not be unreasonable. Rather than choosing an α , Calinski and Corsten (1985) mention the possibility of computing at step s the probability

$$P_s = \Pr \left[Q_{t, t(r-1)} \geq R_s \left[\frac{\text{MS}(E)}{r} \right]^{-1/2} \right],$$

that is, the smallest significance level at which the observed maximum range R_s would lead to the rejection of the null hypothesis associated with the partition at step s . These probabilities can be obtained by using the computer program given by Dunlap, Powell, and Konnerth (1977).

7.7 EXAMPLES USING SAS®

EXAMPLE 7.1: Consider the experiment described in Section 7.2 with $t=5$ treatments and $r=2$ replications in a CRD. The data are given in Table 7.5a.

We use SAS PROC GLM to evaluate the following orthogonal comparisons from Section 7.2.1: (i), (ii), (iii), (iv), using contrast and estimate statements (see Table 7.5a). The contrast statements are used to obtain the contrast SS. The estimate statements provide estimates of the contrasts and their standard error. We also perform Tukey's multiple comparison procedure, providing tests for simple treatment comparisons and simultaneous confidence intervals. We note that generally we would not consider orthogonal contrasts and multiple comparisons for the same experiment.

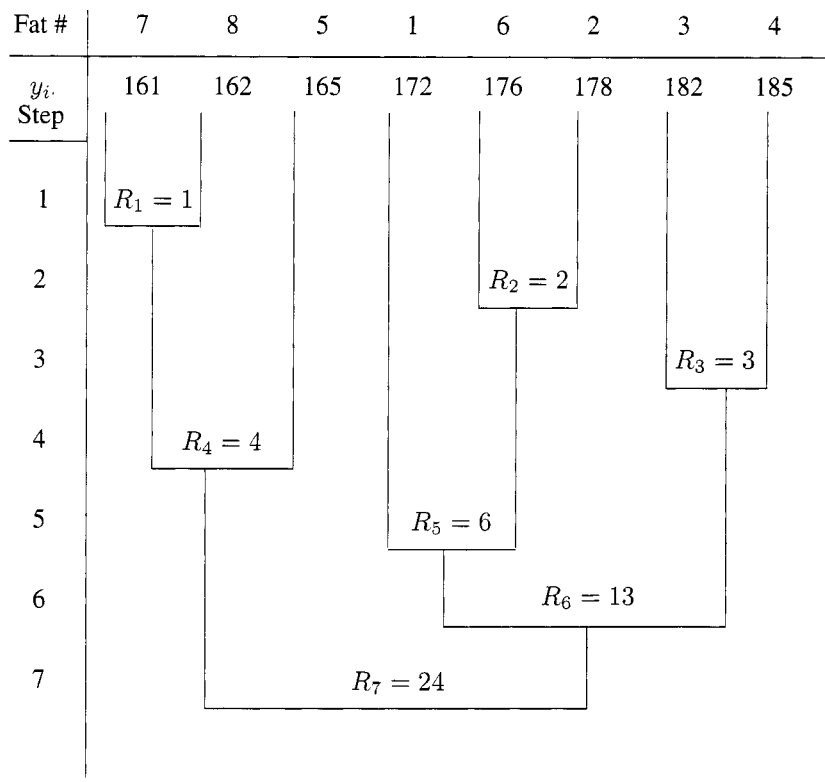


Figure 7.1 Grouping of Snedecor data using the Calinski-Corsten procedure.

One word of caution for the input of the contrast and estimate statements: Since we have used alpha-numeric labeling for the treatments, that is, C, A1, A2, B1, B2, SAS writes them in alphabetical order as A1, A2, B1, B2, C. This requires us to enter the contrast coefficients in this order.

We now turn to the output in Table 7.5b and make the following comments:

- (i) The basic ANOVA provides $\hat{\sigma}_e^2 = .514$.
- (ii) There are significant differences among the treatments ($P = 0.0012$).
- (iii) Writing the treatment LS means in increasing order and using $\alpha = .05$ the results from the Tukey multiple comparison procedure can be summarized as follows, where treatments not connected by the same line are significantly different from each other:

Treatment:	A2	C	B2	A1	B1
each other: LS mean:	13.15	13.35	15.90	18.20	19.10
Tukey ($\alpha = .05$):					

The table in the SAS output provides exact P -values for the comparisons.

- (iv) The simultaneous 95 % confidence intervals for the pairwise comparisons confirm the results given in (iii), but provide additional information about the differences.
- (v) The sum of the contrast SS equals, of course, the SS(Treatments).
- (vi) All specified contrasts are significantly different from zero. □

EXAMPLE 7.2: Kuehl (1994) describes an experiment studying the relationship between grain production and plant density. Using a CRD $t = 5$ plant densities (10, 20, 30, 40, 50) were used, each density was replicated $r = 3$ times. The data are given in Table 7.6a.

Since we have quantitative treatments we use the method of orthogonal polynomials (Section 7.4.2) to obtain the functional relationship (response curve) between yield and plant density. The input statements for SAS PROC GLM are given in Table 7.6b.

We make a few comments:

- (i) Among the input statements we have included “contrast” and “estimate” statements. With the estimate statements we have given the divisor $\sum_i [P_l(z_i)]^2$ of (7.22).
- (ii) The contrast coefficients are obtained from Table 7.3 for $t = 5$.
- (iii) The output shows that the linear and quadratic coefficients are significant ($P, .0001$).
- (iv) The relationship between yield and density can therefore be expressed as

$$\bar{y}_{\cdot i} = \hat{\gamma}_0 P_0(z_i) + \hat{\gamma}_1 P_1(z_i) + \hat{\gamma}_2 P_2(z_i)$$

with

$$\hat{\gamma}_0 = 16.40 = \text{mean}, \quad \hat{\gamma}_1 = 1.19, \quad \hat{\gamma}_2 = -1.01.$$

□

Table 7.5 CRD with Orthogonal Contrasts and Multiple Comparisons

a) Input statements:

```
data pest;
input trt S yield @@;
datalines;
C 12.8 C 13.9
A1 18.5 A1 17.9
A2 12.3 A2 14.0
B1 19.5 B1 18.7
B2 16.0 B2 15.8
;
run;
```

Table 7.5 (Continued)

```
proc glm data=pest;
class trt;
model yield=trt;
lsmeans trt/pdiff adjust=tukey cl;
contrast 'C vs trt' trt 1 1 1 1 -4;
estimate 'C vs trt' trt 1 1 1 1 -4/divisor=4;
contrast 'A vs B' trt 1 1 -1 -1 0;
estimate 'A vs B' trt 1 1 -1 -1 0/divisor=2;
contrast 'A1 vs A2' trt 1 -1 0 0 0;
estimate 'A1 vs A2' trt 1 -1 0 0 0;
contrast 'B1 vs B2' trt 0 0 1 -1 0;
estimate 'B1 vs B2' trt 0 0 1 -1 0;
title1 'COMPLETELY RANDOMIZED DESIGN (t=5, r=2)';
title2 'ORTHOGONAL CONTRASTS AND MULTIPLE COMPARISONS';
run;
```

b) Output:

COMPLETELY RANDOMIZED DESIGN (t=5, r=2)						
ORTHOGONAL CONTRASTS AND MULTIPLE COMPARISONS						
The GLM Procedure						
Class Level Information						
Class	Levels		Values			
trt	5		A1	A2	B1	B2 C
Number of Observations Read				10		
Number of Observations Used				10		
ORTHOGONAL CONTRASTS AND MULTIPLE COMPARISONS						
The GLM Procedure						
Dependent Variable: yield						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	4	59.17400000	14.79350000	28.78	0.0012	
Error	5	2.57000000	0.51400000			
Corrected Total	9	61.74400000				
	R-Square	Coeff Var	Root MSE	yield Mean		
	0.958377	4.497729	0.716938	15.94000		
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
trt	4	59.17400000	14.79350000	28.78	0.0012	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
trt	4	59.17400000	14.79350000	28.78	0.0012	
Least Squares Means						
Adjustment for Multiple Comparisons: Tukey						
	LSMEAN					
trt	yield LSMEAN	Number				
A1	18.2000000	1				
A2	13.1500000	2				

Table 7.5 (Continued)

		B1	19.1000000	3	
		B2	15.9000000	4	
		C	13.3500000	5	
Least Squares Means for effect trt					
Pr > t for H0: LSMean(i)=LSMean(j)					
Dependent Variable: yield					
i/j	1	2	3	4	5
1		0.0047	0.7260	0.1095	0.0057
2	0.0047		0.0022	0.0590	0.9982
3	0.7260	0.0022		0.0332	0.0026
4	0.1095	0.0590	0.0332		0.0773
5	0.0057	0.9982	0.0026	0.0773	
trt yield LSMEAN 95% Confidence Limits					
	A1	18.200000	16.896839	19.503161	
	A2	13.150000	11.846839	14.453161	
	B1	19.100000	17.796839	20.403161	
	B2	15.900000	14.596839	17.203161	
	C	13.350000	12.046839	14.653161	
Least Squares Means for Effect trt					
		Difference	Simultaneous 95%		
		Between	Confidence Limits for		
i	j	Means	LSMean(i)-LSMean(j)		
1	2	5.050000	2.174000	7.926000	
1	3	-0.900000	-3.776000	1.976000	
1	4	2.300000	-0.576000	5.176000	
1	5	4.850000	1.974000	7.726000	
2	3	-5.950000	-8.826000	-3.074000	
2	4	-2.750000	-5.626000	0.126000	
2	5	-0.200000	-3.076000	2.676000	
3	4	3.200000	0.324000	6.076000	
3	5	5.750000	2.874000	8.626000	
4	5	2.550000	-0.326000	5.426000	
Dependent Variable: yield					
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
C vs trt	1	16.77025000	16.77025000	32.63	0.0023
A vs B	1	6.66125000	6.66125000	12.96	0.0155
A1 vs A2	1	25.50250000	25.50250000	49.62	0.0009
B1 vs B2	1	10.24000000	10.24000000	19.92	0.0066
Standard					
Parameter	Estimate	Error	t Value	Pr > t	
C vs trt	3.23750000	0.56678920	5.71	0.0023	
A vs B	-1.82500000	0.50695167	-3.60	0.0155	
A1 vs A2	5.05000000	0.71693793	7.04	0.0009	
B1 vs B2	3.20000000	0.71693793	4.46	0.0066	

Table 7.6 CRD with Quantitative Factors: Orthogonal Polynomials

a) Input statements:

```
data density;
input density yield @@;
datalines;
10 12.2 10 11.5 10 12.3
20 16.1 20 15.3 20 16.6
30 18.6 30 20.1 30 18.4
40 17.7 40 19.3 40 17.0
50 17.8 50 16.4 50 16.7
;
run;

proc glm data=density;
model yield = density;
class density;
contrast 'linear' density -2 -1 0 1 2;
estimate 'linear' density -2 -1 0 1 2/divisor=10;
contrast 'quadratic' density 2 -1 -2 -1 2;
estimate 'quadratic' density 2 -1 -2 -1 2/divisor=14;
contrast 'cubic' density -1 2 0 -2 1;
estimate 'cubic' density -1 2 0 -2 1/divisor=10;
contrast 'quartic' density 1 -4 6 -4 1;
estimate 'quartic' density 1 -4 6 -4 1/divisor= 70;
title1 'CRD WITH QUANTITATIVE FACTORS';
title2 'CONTRASTS USING ORTHOGONAL POLYNOMIALS';
run;
```

b) Output:

CRD WITH QUANTITATIVE FACTORS						
CONTRASTS USING ORTHOGONAL POLYNOMIALS						
The GLM Procedure						
Class Level Information						
Class	Levels	Values				
density	5	10	20	30	40	50
Number of Observations Read						15
Number of Observations Used						15
R-Square	Coeff Var	Root MSE			yield Mean	
0.927949	5.040486	0.826640			16.40000	
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
density	4	88.00666667	22.00166667	32.20	<.0001	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
density	4	88.00666667	22.00166667	32.20	<.0001	
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F	
linear	1	42.72133333	42.72133333	62.52	<.0001	
quadratic	1	42.80380952	42.80380952	62.64	<.0001	
cubic	1	0.28033333	0.28033333	0.41	0.5362	
quartic	1	2.20119048	2.20119048	3.22	0.1029	
Parameter	Estimate	Standard Error		t Value	Pr > t	
linear	1.19333333	0.15092309		7.91	<.0001	
quadratic	-1.00952381	0.12755329		-7.91	<.0001	
cubic	0.09666667	0.15092309		0.64	0.5362	
quartic	0.10238095	0.05704356		1.79	0.102	

7.8 EXERCISES

- 7.1** Consider an experiment to investigate the effects of sugar on the length of pea sections grown in tissue culture. A CRD is used with 5 replications for each of the treatments:

T_1 : Control (nothing added)

T_2 : 2% glucose added

T_3 : 3% glucose added

T_4 : 2% fructose added

T_5 : 1% glucose + 1% fructose added

- (i) Obtain a complete set of meaningful orthogonal contrasts and explain what each contrast means.
- (ii) Suppose we obtain the following results:

Treatment	1	2	3	4	5
Mean	70.1	59.3	58.2	58.0	64.1

and the following partial ANOVA table

Source	SS
Treatments	538.66
Error	245.50
Total	784.16

Partition the $SS(T)$ into single d.f. sums of squares for the orthogonal contrasts obtained in (i) and test the hypotheses that each contrast is equal to zero.

- 7.2** Consider a CRD with 5 treatments, 6 replications for each treatment and 4 observations for every experimental unit. Suppose the treatments represent increasing amounts (x_i) of fertilizer applied to a certain crop.

The following (partial) results are obtained:

x_i	0	2	4	6	8
$\bar{y}_{i..}$	4.9	10.0	13.9	15.7	16.3

$$SS(EE) = 50.0, \quad SS(OE) = 60.0$$

Using the method of orthogonal polynomials investigate whether

- (i) the data exhibit linear and quadratic trends;
- (ii) first and second order terms provide an adequate fit to the data.

7.3 Consider the data in Example 7.1. Obtain a grouping of the treatments using the method described in Section 7.6.

7.4 Consider the data in Example 7.1. Perform Dunnett's procedure (Section 7.5.7) comparing A_1 , A_2 , B_1 , B_2 with C .

7.5 Using the results from Example 7.2 obtain the prediction equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \text{ density} + \hat{\beta}_2 (\text{density})^2$$

using (i) the form of the $P_l(z_i)$ given in Section 7.4.2 and (ii) fitting the second degree polynomial directly.

This Page Intentionally Left Blank

CHAPTER 8

Use of Supplementary Information

8.1 INTRODUCTION

As we have pointed out earlier, one of the main purposes of experimental design is the reduction of error. One important component of the overall error is the unit error, exemplified by σ_u^2 (see Section 6.3), expressing a certain amount of heterogeneity among the EUs. Generally speaking, such variability among the EUs may be systematic or random. Consider the following examples for systematic variation:

- (i) a fertility trend exists in a piece of land used for an agronomic trial;
- (ii) animals used in a pharmaceutical experiment may come from different litters;
- (iii) the experimental material for an industrial experiment may come from different production processes;

and for random variation:

- (iv) plants for a growth trial may be of different heights at the beginning of the trial;
- (v) animals for a dietary study may have different initial weights;
- (vi) individuals for an educational study may have different abilities as documented by I.Q. or earlier test scores.

In the case of systematic variation, knowledge of the underlying reasons will lead to blocking and hence to more complex designs which will be discussed in subsequent chapters. For random variation the additional (supplementary) information can, under certain conditions, be used effectively to reduce the error in a CRD. This procedure, introduced by Fisher (1932) and referred to as *analysis of covariance*, is the topic of this chapter. For a general description see also Cochran (1957).

8.2 MOTIVATION OF THE PROCEDURE

In its simplest form we have two measurements for each EU (assuming that the EU and OU are identical):

- y : the response to the treatment
 x : the supplementary information.

Examples for the cases (iv)–(vi) above might be

- (iv) y denotes the growth of the plant after exposure to a treatment, x denotes the initial height;
- (v) y denotes the final weight after the treatment, x denotes the initial weight;
- (vi) y denotes the test score after the treatment, x denotes a test score before the treatment.

It is assumed that the supplementary information, or covariate, x is independent of the treatment. This is a rather crucial assumption with respect to the unbiased estimation of differences among treatment effects (Rosenbaum, 1984). This means that covariates must be either obtained before the treatment assignment and/or application, or they must be known not to be influenced by the treatment, for example the outside temperature during a physical exercise in a clinical study. Also, it is known or suspected that there exists a functional relationship between the response y and the covariate x . We emphasize that this relationship may not be a causal relationship. The covariate x may be correlated with something, often unknown, which causes extraneous variation in the response y (Smith, 1957). To illustrate this point Smith (1957) describes an example where in a field experiment x represents the amount of weed present in a field plot. The weed itself may not have affected the crop yield, rather it may have been a surrogate for the soil acidity present in the plot, which is correlated with the growth of weed as well as the crop under investigation. For purposes of our present discussion we shall assume that the relationship between x and y is linear. In its simplest form, in the absence of treatment effects, the data may look as given in Figure 8.1.

It is informative, for purposes of illustration and motivation, to consider the model (6.3)

$$T_{ik} = T_i + U_k$$

and write it in the form

$$T_{ik} = T_i + \alpha + \beta(X_k - \bar{X}.) + U_k^*, \quad (8.1)$$

that is, the unit contribution U_k is modeled as

$$U_k = \alpha + \beta(X_k - \bar{X}.) + U_k^*,$$

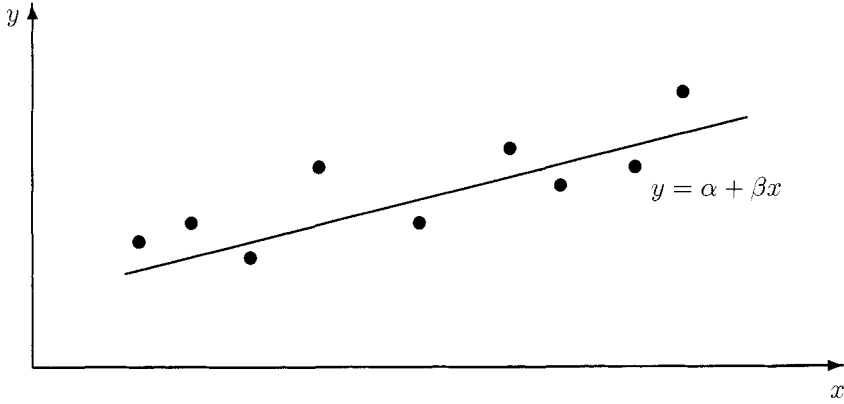


Figure 8.1 Relationship between covariate and response.

where α, β are constants and X_k is the value of a covariate for EU k . Model (6.6) can then be rewritten as

$$\begin{aligned} y_{ij} &= \mu + \tau_i + \alpha + \beta \sum_k \delta_{ij}^k (X_k - \bar{X}.) + \sum_k \delta_{ij}^k u_k^* \\ &= \mu^* + \tau_i + \beta(x_{ij} - \bar{x}..) + \omega_{ij}^*. \end{aligned} \quad (8.2)$$

In the absence of treatment effects, (8.2) reduces to

$$y_{ij} = \mu^* + \beta(x_{ij} - \bar{x}..) + \omega_{ij}^* \quad (8.3)$$

or

$$z_{ij} = y_{ij} - \beta(x_{ij} - \bar{x}..) = \mu^* + \omega_{ij}^*. \quad (8.4)$$

The form of (8.4) suggests that if we adjust the observations y_{ij} by the concomitant information $\beta(x_{ij} - \bar{x}..)$, then the new “observations” z_{ij} are constant apart from noise ω_{ij}^* , where

$$E_R(\omega_{ij}^*) = 0$$

and

$$\text{var}_R(\omega_{ij}^*) = \left(1 - \frac{1}{N}\right) \sigma_{u^*}^2$$

and, most importantly,

$$\sigma_{u^*}^2 < \sigma_u^2.$$

In fact, $\sigma_{u^*}^2$ measures the variability of the EUs around the regression line. It seems then natural to obtain the adjusted observations (8.4) and perform the usual analysis on them. The problem, of course, is that usually we either do not know β or we think we know β but it is not the correct value (in many cases $\beta = 1$ is used such as $z = \text{post-test-pretest scores}$, or $z = \text{final weight-initial weight}$). The question then arises: How do we use the supplementary information and how do we make adjustments?

8.3 ANALYSIS OF COVARIANCE PROCEDURE

8.3.1 Basic Model

Generalizing model (8.3), we now consider the model

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + e_{ij}^* \quad (8.5)$$

for the observations from a CRD where supplementary information of the form described above is available ($i = 1, 2, \dots, t; j = 1, 2, \dots, r$). The e_{ij}^* are considered to be i.i.d. random variables with mean zero and variance $\sigma_{e^*}^2$ (to relate e_{ij}^* to our earlier discussion we can think of e_{ij}^* being of the form

$$e_{ij}^* = \omega_{ij}^* + \nu_{ij} + \eta_{ij} = \varepsilon_{ij}^* + \eta_{ij} \quad (8.6)$$

with $\varepsilon_{ij}^* \sim (0, \sigma_{\varepsilon^*}^2)$ and $\sigma_{e^*}^2 = \sigma_{\varepsilon^*}^2 + \sigma_{\eta}^2$. A graph of the data (y_{ij}, x_{ij}) may then be as illustrated in Figure 8.2 for $t = 3, r = 6$, where the lines labeled T_1, T_2, T_3 represent the linear relationship between y and x for treatments 1, 2, 3, respectively, that is, (8.5) for $i = 1, 2, 3$.

Our aim is to estimate contrasts among the treatment effects, that is, $\sum c_i \tau_i$, and test hypotheses about treatment effects, such as $H_0: \tau_1 = \tau_2 = \dots = \tau_t = 0$. Assuming again for a moment that we know β , it follows easily from (8.5) and Figure 8.2 that, for example,

$$\begin{aligned} \hat{\tau}_1 - \hat{\tau}_2 &= \bar{y}_{1.} - \bar{y}_{2.} - \beta[(\bar{x}_{1.} - \bar{x}_{..}) - (\bar{x}_{2.} - \bar{x}_{..})] \\ &= y_{A_1} - y_{A_2}, \end{aligned} \quad (8.7)$$

where y_{A_i} is the y -value at $x = \bar{x}_{..}$ for treatment i . The estimator (8.7) is, of course, the corresponding difference between the treatment means plus an adjustment due to differences in the covariates for the two treatments. For that reason $y_{A_1} - y_{A_2}$ is referred to as the *adjusted treatment difference*.

8.3.2 Least Squares Analysis

Let us now turn to the more important case with β unknown. We shall use the method of least squares (for the arithmetic of analysis of covariance see also Section 4.13) to obtain estimates of estimable functions involving μ , τ_i , and β . Using model (8.5) we obtain the normal equations (NE) by minimizing the expression

$$Q = \sum_{i,j} [y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..})]^2 \quad (8.8)$$

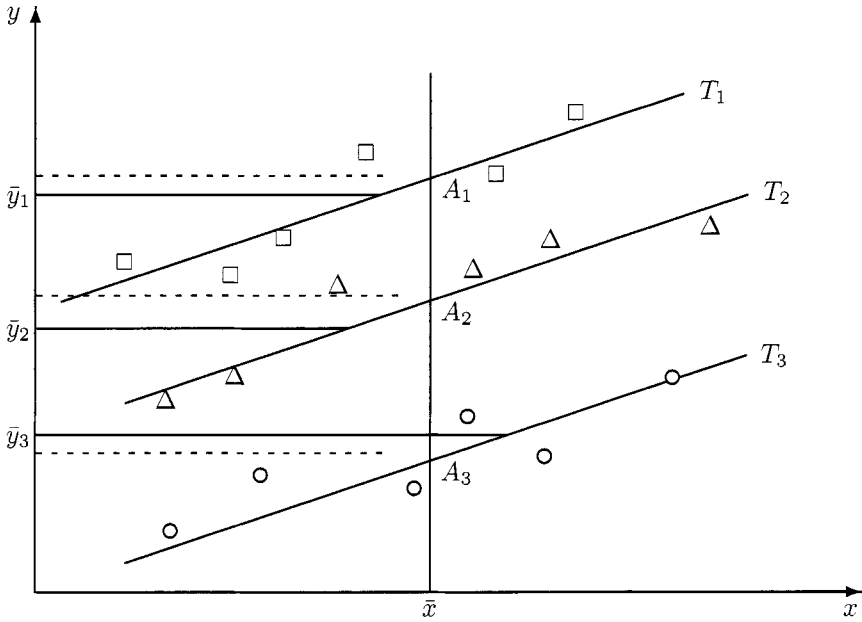


Figure 8.2 Graph of data (y, x) from CRD.

with respect to μ , $\tau_i (i = 1, 2, \dots, t)$, β . Differentiating (8.8) w.r.t. these parameters and equating the derivatives to zero leads to the NE

$$tr\hat{\mu} + r \sum_i \hat{\tau}_i + \hat{\beta} \sum_{ij} (x_{ij} - \bar{x}_{..}) = y_{..} \quad (8.9)$$

$$r\hat{\mu} + r\hat{\tau}_i + \hat{\beta} \sum_j (x_{ij} - \bar{x}_{..}) = y_{i.} \quad (8.10)$$

$$(i = 1, 2, \dots, t)$$

$$\sum_{ij} (x_{ij} - \bar{x}_{..})\hat{\mu} + \sum_{ij} (x_{ij} - \bar{x}_{..})\hat{\tau}_i + \hat{\beta} \sum_{ij} (x_{ij} - \bar{x}_{..})^2 = \sum_{ij} y_{ij}(x_{ij} - \bar{x}_{..}), \quad (8.11)$$

where $y_{i.} = \sum_j y_{ij}$, $y_{..} = \sum_{ij} y_{ij}$. With $\sum_{ij} (x_{ij} - \bar{x}_{..}) = 0$ and putting $\sum_i \hat{\tau}_i = 0$ (since $\sum \tau_i = 0$) equations (8.9)–(8.11) can be simplified. From (8.9) we obtain

$$\hat{\mu} = \bar{y}_{..} \quad (8.12)$$

and from (8.10) and (8.12) we obtain

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}). \quad (8.13)$$

Substituting (8.12) and (8.13) into (8.11) yields

$$\sum_{ij} (x_{ij} - \bar{x}_{..})[(\bar{y}_{i.} - \bar{y}_{..}) - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})] + \hat{\beta} \sum_{ij} (x_{ij} - \bar{x}_{..})^2 = \sum_{ij} y_{ij}(x_{ij} - \bar{x}_{..})$$

Table 8.1 Auxiliary ANOVA for CRD

Source	SS(y)	SS(x)	SP(x, y)
Treatments	T_{yy}	T_{xx}	T_{xy}
Error	E_{yy}	E_{xx}	E_{xy}
Total	S_{yy}	S_{xx}	S_{xy}

or

$$\begin{aligned}
 r \sum_i (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}) + \hat{\beta} \left[\sum_{ij} (x_{ij} - \bar{x}_{..})^2 - r \sum_i (\bar{x}_{i.} - \bar{x}_{..})^2 \right] \\
 = \sum_{ij} (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}). \quad (8.14)
 \end{aligned}$$

It is useful to introduce some simplifying notation. In Table 8.1 we give symbols for the various sums of squares and sums of products for a CRD, using the y_{ij} and x_{ij} as “observations.” In Table 8.1, the $SS(y)$ are the same as those in Table 6.1, the $SS(x)$ are obtained analogously with the x_{ij} substituted for the y_{ij} , and the $SP(x, y)$ are sums of products rather than sums of squares, for example,

$$T_{xy} = r \sum_i (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}).$$

Using this notation and using the algebraic fact that $T_{pq} + E_{pq} = S_{pq}$, where p, q are replaced by x and/or y , we rewrite (8.14) as

$$T_{xy} + \hat{\beta}(S_{xx} - T_{xx}) = S_{xy}$$

or

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}}. \quad (8.15)$$

8.3.3 Least Squares Means

Under our model assumptions it follows that $\hat{\mu}$, $\hat{\tau}_i$, and $\hat{\beta}$ given by (8.12), (8.13), and (8.15), respectively, are the BLUEs of μ , τ_i , and β , respectively. Hence

$$\hat{\mu} + \hat{\tau}_i = \bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \quad (8.16)$$

is the BLUE for $\mu + \tau_i$, the response for treatment i . The right-hand side of (8.16) is often referred to as the *adjusted treatment mean*, adjusted for differences in the co-variates. It is also called the *least squares mean* for treatment i , which we shall write,

for future references, as $\text{LSM}(\tau_i)$ [Searle, Speed, and Milliken (1980) refer to $\mu + \tau_i$ as the population marginal mean (PMM) and to $\hat{\mu} + \hat{\tau}_i$ as the estimated PMM; Lane and Nelder (1982) refer to $\hat{\mu} + \hat{\tau}_i$ as the predictive margin]. It follows further that the BLUE for a treatment contrast $\sum c_i \tau_i$ with $\sum c_i = 0$ is given by

$$\sum_i c_i \hat{\tau}_i = \sum_i c_i \bar{y}_{i.} - \hat{\beta} \sum_i c_i \bar{x}_{i.}. \quad (8.17)$$

As a special case we have

$$\widehat{\tau_i - \tau_{i'}} = \hat{\tau}_i - \hat{\tau}_{i'} = \bar{y}_{i.} - \bar{y}_{i'}. - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{i'}.),$$

which is, of course, (8.7) with β replaced by $\hat{\beta}$.

To obtain the variances of the estimators given above we use the fact that $\bar{y}_{i.}$ and $\hat{\beta}$ are uncorrelated. We then find

$$\text{var}(\hat{\beta}) = \frac{\sigma_{e^*}^2}{E_{xx}} \quad (8.18)$$

$$\text{var}(\hat{\mu} + \hat{\tau}_i) = \left[\frac{1}{r} + \frac{(\bar{x}_{i.} - \bar{x}_{..})^2}{E_{xx}} \right] \sigma_{e^*}^2 \quad (8.19)$$

$$\text{cov}(\hat{\mu} + \hat{\tau}_i, \hat{\mu} + \hat{\tau}_{i'}) = \frac{(\bar{x}_{i.} - \bar{x}_{..})(\bar{x}_{i'.} - \bar{x}_{..})}{E_{xx}} \sigma_{e^*}^2 \quad (i \neq i') \quad (8.20)$$

$$\text{var} \left(\sum_i c_i \hat{\tau}_i \right) = \left[\frac{\sum_i c_i^2}{r} + \frac{(\sum_i c_i \bar{x}_{i.})^2}{E_{xx}} \right] \sigma_{e^*}^2 \quad (8.21)$$

with $\sum c_i = 0$.

8.3.4 Formulation in Matrix Notation

In order to estimate $\sigma_{e^*}^2$ and to test $H_0: \tau_1 = \tau_2 = \dots = \tau_t$ and $H_0: \beta = 0$ we turn to the analysis of variance. For the derivation of the ANOVA table we make use of results in Sections 4.12.2 and 4.12.3 where we have discussed the general case. It is, therefore, useful to reformulate some of the results above in matrix notation.

We write model (8.5) as

$$\mathbf{y} = \mathbf{J} \mu + \mathbf{X}_\tau \boldsymbol{\tau} + \mathbf{X}_\beta \beta + \mathbf{e}^*, \quad (8.22)$$

where $\mathbf{y} = (y_{11}, y_{12}, \dots, y_{tr})'$ is a $tr \times 1$ column vector of the observations, \mathbf{J} is a $tr \times 1$ column vector of unity elements, \mathbf{X}_τ is a $tr \times t$ matrix of known constants (zero or one), $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_t)'$ is a $t \times 1$ vector of treatment effects, $\mathbf{X}_\beta = (x_{11} - \bar{x}_{..}, x_{12} - \bar{x}_{..}, \dots, x_{tr} - \bar{x}_{..})'$ is a $tr \times 1$ vector of the covariates (expressed as deviation from the mean), and $\mathbf{e}^* = (e_{11}^*, e_{12}^*, \dots, e_{tr}^*)'$ is a $tr \times 1$ vector of errors. The NE are then of the form

$$\begin{bmatrix} \mathbf{J}' \mathbf{J} & \mathbf{J}' \mathbf{X}_\tau & \mathbf{J}' \mathbf{X}_\beta \\ \mathbf{X}_\tau' \mathbf{J} & \mathbf{X}_\tau' \mathbf{X}_\tau & \mathbf{X}_\tau' \mathbf{X}_\beta \\ \mathbf{X}_\beta' \mathbf{J} & \mathbf{X}_\beta' \mathbf{X}_\tau & \mathbf{X}_\beta' \mathbf{X}_\beta \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\tau}} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{J}' \mathbf{y} \\ \mathbf{X}_\tau' \mathbf{y} \\ \mathbf{X}_\beta' \mathbf{y} \end{bmatrix}, \quad (8.23)$$

which can be rewritten as

$$\begin{bmatrix} tr & r\mathbf{J}'_t & 0 \\ r\mathbf{J}_t & r\mathbf{I}_t & \mathbf{X}'_\tau \mathbf{X}_\beta \\ 0 & \mathbf{X}'_\beta \mathbf{X}_\tau & \mathbf{X}'_\beta \mathbf{X}_\beta \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} y_{..} \\ \mathbf{X}'_\tau \mathbf{y} \\ \mathbf{X}'_\beta \mathbf{y} \end{bmatrix}. \quad (8.24)$$

With $\mathbf{J}'_t \hat{\tau} = 0$ it follows from (8.23) and (8.24) that

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..} \\ \hat{\tau} &= \frac{1}{r} \mathbf{X}'_\tau (\mathbf{y} - \mathbf{J} \bar{y}_{..} - \mathbf{X}_\beta \hat{\beta}) \\ \hat{\beta} &= (\mathbf{X}'_\beta \mathbf{X}_\beta)^{-1} \mathbf{X}'_\beta (\mathbf{y} - \mathbf{J} \bar{y}_{..} - \mathbf{X}_\tau \hat{\tau}). \end{aligned}$$

Evaluating these expressions leads, of course, to the estimators (8.12), (8.13), and (8.15).

8.3.5 ANOVA Table

We now turn to the ANOVA table. Since this is a nonorthogonal ANOVA, we use the notation established in Chapter 4 to indicate how the various sums of squares have been obtained. Specifically, we obtain the treatment SS as

$$SS(\mathbf{X}_\tau | \mathbf{J}, \mathbf{X}_\beta) = SS(\mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta) - SS(\mathbf{J}, \mathbf{X}_\beta) \quad (8.25)$$

with

$$\begin{aligned} SS(\mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta) &= \hat{\mu} y_{..} + \hat{\tau}' \mathbf{X}'_\tau \mathbf{y} + \hat{\beta}' \mathbf{X}'_\beta \mathbf{y} \\ &= \hat{\mu} y_{..} + \sum_i \hat{\tau}_i y_{i.} + \hat{\beta}' \sum_{ij} (x_{ij} - \bar{x}_{..}) y_{ij} \\ &= tr \bar{y}^2 + r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &\quad + \frac{E_{xy}}{E_{xx}} \left[\sum_{ij} (x_{ij} - \bar{x}_{..}) (y_{ij} - \bar{y}_{..}) - r \sum_i (\bar{x}_{i.} - \bar{x}_{..}) (\bar{y}_{i.} - \bar{y}_{..}) \right] \\ &= tr \bar{y}^2 + T_{yy} + \frac{E_{xy}^2}{E_{xx}} \end{aligned} \quad (8.26)$$

using the notation of Table 8.1. To obtain $SS(\mathbf{J}, \mathbf{X}_\beta)$ we use the model

$$\mathbf{y} = \mathbf{J}\mu + \mathbf{X}_\beta \beta + \mathbf{e}^*,$$

which leads to the NE

$$\begin{bmatrix} \mathbf{J}'\mathbf{J} & \mathbf{J}'\mathbf{X}_\beta \\ \mathbf{X}'_\beta \mathbf{J} & \mathbf{X}'_\beta \mathbf{X}_\beta \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} y_{..} \\ \mathbf{X}'_\beta \mathbf{y} \end{bmatrix}$$

and the familiar solutions

$$\begin{aligned}\tilde{\mu} &= \bar{y}_{..} \\ \tilde{\beta} &= (\mathbf{X}'_{\beta} \mathbf{X}_{\beta})^{-1} \mathbf{X}'_{\beta} \mathbf{y} \\ &= \frac{\sum_{ij} (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})}{\sum_{ij} (x_{ij} - \bar{x}_{..})^2} \\ &= \frac{S_{xy}}{S_{xx}}\end{aligned}$$

using again the notation from Table 8.1. It follows then that

$$SS(\mathcal{J}, \mathbf{X}_{\beta}) = \tilde{\mu} y_{..} + \tilde{\beta} \mathbf{X}'_{\beta} \mathbf{y} = tr \bar{y}^2 + \frac{S_{xy}^2}{S_{xx}}. \quad (8.27)$$

Substituting (8.26) and (8.27) into (8.25), we obtain

$$SS(\mathbf{X}_{\tau} | \mathcal{J}, \mathbf{X}_{\beta}) = T_{yy} - \frac{S_{xy}^2}{S_{xx}} + \frac{E_{xy}^2}{E_{xx}}. \quad (8.28)$$

We proceed in a similar fashion to obtain the regression SS as

$$SS(\mathbf{X}_{\beta} | \mathcal{J}, \mathbf{X}_{\tau}) = SS(\mathcal{J}, \mathbf{X}_{\tau}, \mathbf{X}_{\beta}) - SS(\mathcal{J}, \mathbf{X}_{\tau}). \quad (8.29)$$

In order to obtain $SS(\mathcal{J}, \mathbf{X}_{\tau})$, we use the model

$$\mathbf{y} = \mathcal{J}\mu + \mathbf{X}_{\tau}\boldsymbol{\tau} + \mathbf{e}^*,$$

which leads to the NE

$$\begin{bmatrix} \mathcal{J}'\mathcal{J} & \mathcal{J}'\mathbf{X}_{\tau} \\ \mathbf{X}'_{\tau}\mathcal{J} & \mathbf{X}'_{\tau}\mathbf{X}_{\tau} \end{bmatrix} \begin{bmatrix} \tilde{\mu} \\ \tilde{\boldsymbol{\tau}} \end{bmatrix} = \begin{bmatrix} y_{..} \\ \mathbf{X}'_{\tau}\mathbf{y} \end{bmatrix}$$

and the solutions

$$\begin{aligned}\tilde{\mu} &= \bar{y}_{..} \\ \tilde{\boldsymbol{\tau}} &= (\mathbf{X}'_{\tau} \mathbf{X}_{\tau})^{-1} [\mathbf{X}'_{\tau} \mathbf{y} - \mathbf{X}'_{\tau} \mathcal{J} \tilde{\mu}] \\ &= \frac{1}{r} \mathbf{X}'_{\tau} (\mathbf{y} - \mathcal{J} \bar{y}_{..}) \\ &= \begin{bmatrix} \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{t.} - \bar{y}_{..} \end{bmatrix}.\end{aligned}$$

Hence,

$$\begin{aligned}
 SS(\mathbf{J}, \mathbf{X}_\tau) &= \tilde{\mu}y_{..} + \tilde{\tau}'\mathbf{X}'_\tau\mathbf{y} \\
 &= tr\bar{y}^2 + \frac{1}{r}(\mathbf{y} - \mathbf{J}\bar{y}_{..})'\mathbf{X}_\tau\mathbf{X}'_\tau\mathbf{y} \\
 &= tr\bar{y}^2 + r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \\
 &= tr\bar{y}^2 + T_{yy}.
 \end{aligned} \tag{8.30}$$

Substituting (8.26) and (8.30) into (8.29), we obtain

$$SS(\mathbf{X}_\beta|\mathbf{J}, \mathbf{X}_\tau) = \frac{E_{xy}^2}{E_{xx}}. \tag{8.31}$$

Finally, the error sum of squares is obtained as

$$\begin{aligned}
 SS(\mathbf{I}|\mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta) &= SS(\text{Total})_{\text{unadj.}} - SS(\mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta) \\
 &= \sum_{ij} y_{ij}^2 - tr\bar{y}^2 - T_{yy} - \frac{E_{xy}^2}{E_{xx}} \\
 &= E_{yy} - \frac{E_{xy}^2}{E_{xx}}.
 \end{aligned} \tag{8.32}$$

We point out, in passing, the similarity between (8.32) and the form of the error sum of squares for a simple linear regression model, such as (8.3), which is our notation is given by $S_{yy} - S_{xy}^2/S_{xx}$.

The complete ANOVA table is given in Table 8.2.

It follows from $E(\text{MS})$ in Table 8.2 that an estimator for $\sigma_{e^*}^2$ is given by

$$\begin{aligned}
 \hat{\sigma}_{e^*}^2 &= MS(\mathbf{I}|\mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta) \\
 &= \left[E_{yy} - \frac{E_{xy}^2}{E_{xx}} \right] / [t(r-1) - 1].
 \end{aligned} \tag{8.33}$$

The form of (8.33) shows explicitly that, unless there is no or only a weak linear relationship between the observation y and the covariate x , the variance estimator $\hat{\sigma}_{e^*}^2$ is smaller than the comparable variance estimator $\hat{\sigma}_e^2$ of (6.29), that is, for the CRD without supplemental information (see also Section 8.5.1).

The form of the $E(\text{MS})$ in Table 8.2 suggests immediately to test $H_0: \tau_1 = \tau_2 = \dots = \tau_t = 0$ by the F -test, as an approximation to the randomization test (Robinson, 1973) for large r ,

$$F = \frac{MS(\mathbf{X}_\tau|\mathbf{J}, \mathbf{X}_\beta)}{MS(\mathbf{I}|\mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta)}$$

and $H_0: \beta = 0$ by

$$F = \frac{MS(\mathbf{X}_\beta|\mathbf{J}, \mathbf{X}_\tau)}{MS(\mathbf{I}|\mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta)}$$

Table 8.2 Analysis of Covariance Table

Source	d.f.	SS	E(MS)	
$\mathbf{X}_\tau \mathcal{J}, \mathbf{X}_\beta$	$t - 1$	$T_{yy} - \frac{S_{xy}^2}{S_{xx}}$	$\sum \tau_i^2 - 1$	$\frac{r^2 \left(\sum \tau_i \bar{x}_i \right)^2}{(t - 1)(E_{xx} + T_{xx})}$
$\mathbf{X}_\beta \mathcal{J}, \mathbf{X}_\tau$	1	$\frac{E_{xy}^2}{E_{xx}}$	$\sigma_{e^*}^2 + \beta^2 E_{xx}$	
$\mathbf{I} \mathcal{J}, \mathbf{X}_\tau, \mathbf{X}_\beta$	$t(r - 1) - 1$	$E_{yy} - \frac{F_{xy}^2}{F_{xx}}$	σ_c^2	
Total	$tr - 1$			

using the appropriate d.f. as indicated in Table 8.2. For a numerical example see Section 8.8.

8.4 TREATMENT COMPARISONS

8.4.1 Preplanned Comparisons

As discussed in Chapter 7, the overall null hypothesis $\tau_1 = \tau_2 = \cdots = \tau_t = 0$ is often less important and less informative than specific hypotheses of the form $\sum_i c_i \tau_i = 0$ with $\sum_i c_i = 0$. We have shown in Section 8.3 that

$$\text{var} \left(\sum_i c_i \hat{\tau}_i \right) = \left[\frac{\sum c_i^2}{r} + \frac{\left(\sum c_i \bar{x}_i \right)^2}{E_{xx}} \right] \sigma_e^2.$$

[see (8.21)]

Using (8.21) and (8.33) we can then perform the usual tests on single d.f. contrasts either in the context of the ANOVA by forming

$$\text{SS} \left(\sum c_i \tau_i \right) = \frac{\left(\sum c_i \hat{\tau}_i \right)^2}{\frac{\sum c_i^2}{r} + \frac{\left(\sum c_i \bar{x}_i \right)^2}{E_{xx}}} \quad (8.34)$$

and

$$F = \frac{\text{SS} \left(\sum c_i \tau_i \right)}{\text{MS}(\mathbf{I}\mathcal{J}, \mathbf{X}_\tau, \mathbf{X}_\beta)} \quad (8.35)$$

or, equivalently, by using

$$t = \frac{\sum c_i \hat{\tau}_i}{\hat{\sigma}_e^* \sqrt{\frac{\sum c_i^2}{r} + \frac{\left(\sum c_i \bar{x}_i \right)^2}{E_{xx}}}}. \quad (8.36)$$

(We should mention here that for a complete set of orthogonal contrasts, $\mathbf{C}'\boldsymbol{\tau}$, satisfying (7.7) and (7.8), we no longer have the result (7.6). The reason for this is that even though the contrasts are orthogonal, their estimators are not, that is, they are correlated as follows from (8.20).) In particular, we may be interested in testing hypotheses about simple treatment differences $\tau_i - \tau_{i'}$ and, if appropriate, use the multiple comparison procedures discussed in Chapter 7. Since we no longer can compare treatment means but rather have to use LS means, which not only may have different variances but are also correlated, this leads to certain complications and calls for modifications of the procedures discussed earlier.

8.4.2 Multiple Comparison Procedures

Suppose we want to use Duncan's Multiple Range Test (see Section 7.5.5). We may begin by arranging the LS means in increasing order

$$\hat{\mu} + \hat{\tau}_{[1]}, \hat{\mu} + \hat{\tau}_{[2]}, \dots, \hat{\mu} + \hat{\tau}_{[t]}.$$

Following the established procedure we then compare $\hat{\mu} + \hat{\tau}_{[1]}$ versus $\hat{\mu} + \hat{\tau}_{[t]}$, that is, $\hat{\tau}_{[1]}$ versus $\hat{\tau}_{[t]}$, by considering

$$\hat{\tau}_{[t]} - \hat{\tau}_{[1]} = \bar{y}_{[t]} - \bar{y}_{[1]} - \hat{\beta}(\bar{x}_{[t]} - \bar{x}_{[1]}), \quad (8.37)$$

where $\bar{x}_{[t]}$ is the mean of the x -variates corresponding to $y_{[t]}$, and so forth, and comparing it with

$$Q_{\alpha_t, t, \nu} \sqrt{\frac{1}{2} \left[\frac{2}{r} + \frac{(\bar{x}_{[t]} - \bar{x}_{[1]})^2}{E_{xx}} \right]} \hat{\sigma}_{e^*}, \quad (8.38)$$

where $\nu = t(r - 1) - 1$ are the d.f. for error (see Kramer, 1957; Miller, 1981). If (8.37) is larger than (8.38), then the effects of the corresponding treatments are judged to be different; if (8.37) is smaller than (8.38) then the treatments are judged to be not different from each other, and ordinarily the comparison procedure would stop (see Chapter 7). Such a property no longer holds in the present situation where the variance of the difference of LS means is not constant but does depend on the x -means for the treatments involved [see (8.38)]. For this reason treatments are compared with different precisions, more precise if the x -means are close together, less precise if the x -means are far apart. Hence the nonsignificance of $\hat{\tau}_{[t]} - \hat{\tau}_{[1]}$ may be due to the fact that, just by chance, $\bar{x}_{[t]} - \bar{x}_{[1]}$ is rather large. It can very well happen then that two other LS means, say $\text{LSM}(\tau_{[i]})$ and $\text{LSM}(\tau_{[i']})$, within the remaining set are significantly different from each other simply because their x -means are close together, that is, the quantity

$$Q_{\alpha_k, k, \nu} \sqrt{\frac{1}{2} \left[\frac{2}{r} + \frac{(\bar{x}_{[i]} - \bar{x}_{[i']})^2}{E_{xx}} \right]} \hat{\sigma}_{e^*}$$

(with $k < t$) is not only smaller than (8.38) but also smaller than $\hat{\tau}_{[i]} - \hat{\tau}_{[i']}$. The important point of this discussion is that even after a nonsignificant result for LS means of range l , we may have to continue comparing LS means of range less than l . This may indeed lead to making all possible $t(t - 1)/2$ comparisons.

The procedure just described may be rather tedious. An alternative, much simpler but possibly less satisfactory, method is to make all comparisons by using a constant variance, namely the average variance of all simple treatment comparisons. We find that

$$\begin{aligned} \text{av. var}(\hat{\tau}_i - \hat{\tau}_{i'}) &= \frac{1}{t(t-1)} \sum_{i \neq i'} \left[\frac{2}{r} + \frac{(\bar{x}_i - \bar{x}_{i'})^2}{E_{xx}} \right] \sigma_{e^*}^2 \\ &= \frac{2}{r} \left[1 + \frac{1}{t-1} \frac{T_{xx}}{E_{xx}} \right] \sigma_{e^*}^2. \end{aligned} \quad (8.39)$$

Since we assume that the x -values are not affected by the treatments, we would have that, on average,

$$\frac{T_{xx}}{t-1} = \frac{E_{xx}}{t(r-1)}, \quad (8.40)$$

that is, the treatment mean square and the error mean square for the covariate are equal. If we use (8.40) in (8.39), the expression for the average variance reduces to

$$\text{av. var}(\hat{\tau}_i - \hat{\tau}_{i'}) = \frac{2}{r} \left[1 + \frac{1}{t(r-1)} \right] \sigma_e^2 \quad (8.41)$$

independent of the actually observed x -values. (A slightly different result was obtained by Cox (1957) by considering the covariates as normally distributed random variables). We may then use any of the multiple comparison procedures with (8.40) or (8.41). By doing so we must, however, realize that this procedure favors certain comparisons, those that have a variance larger than (8.41), over others, those that have a variance less than (8.41). Hence, care should be used in interpreting the results. Obviously, this procedure works quite well if the x -means are not too different from each other. This is an ideal situation, one that has also been advocated by Cox (1982) for purposes of randomization analysis.

Arguments similar to those above can be made for other multiple comparison procedures, extending, for example, the Tukey procedure to the Tukey-Kramer procedure based on the result given by Kramer (1957). An example will be given in Section 8.8. For more details the reader is referred to Hochberg and Tamhane (1987).

8.5 VIOLATION OF ASSUMPTIONS

During our discussion so far we have made a number of assumptions, some implicit and some explicit. These assumptions can be summarized as follows:

- (i) There exists a linear relationship between the covariate x and the observation y .
- (ii) The relationship between y and x is the same for each treatment.
- (iii) The covariates are not affected by the treatments.
- (iv) The observations come from a normal distribution.

In this section we shall consider these assumptions, how they may be checked in a given situation, and point to the implication of the violation of these assumptions.

8.5.1 Linear Relationship between x and y

Suppose two random variables x and y have a bivariate normal distribution with means μ_x, μ_y , variances σ_x^2, σ_y^2 , and covariance $\rho\sigma_x\sigma_y$. Then the conditional distribution of y , given x , is normal with mean $\mu_y + \beta(x - \mu_x)$ and variance $\sigma_y^2(1 - \rho^2)$, where $\beta = \rho\sigma_y/\sigma_x$. The variance $\sigma_y^2(1 - \rho^2)$ is the variance of y about the regression line $y = \mu_y + \beta(x - \mu_x)$. In our situation we have t regression lines of the form $y =$

$\mu_{yi} + \beta(x - \mu_{xi})$ ($i = 1, 2, \dots, t$) with $\mu_{yi} = \mu + \tau_i$, and the variance of y about each line is $\sigma_y^2(1 - \rho^2)$ (see Figure 8.2), where in our notation $\sigma_y^2 = \sigma_e^2$. This implies that the increase in precision using the covariate x is given by $(1 - \rho^2)$, or if we account for the fact that β has to be estimated and hence leads to an increase in the variance of comparisons, the increase in precision is on the average given by

$$I = (1 - \rho^2) \frac{t(r-1) - 1}{t(r-1) - 2} \quad (8.42)$$

(Cochran, 1957; Cox and McCullagh, 1982). It follows then that an increase in precision will be realized only if ρ is of a reasonable magnitude. If, for example, $\rho = 0$, that is, there does not exist a linear relationship between x and y , then $I > 1$ and, in fact, information has been lost rather than gained. In order to get some idea how large ρ would have to be for the analysis of covariance to be worthwhile we consider (8.42) in connection with (8.41), that is, we compare the average variance of treatment comparisons using the covariate

$$\text{av. var}(\hat{\tau}_i - \hat{\tau}_{i'}) = \frac{2}{r} \left(1 + \frac{1}{t(r-1)} \right) (1 - \rho^2) \cdot \frac{t(r-1) - 1}{t(r-1) - 2} \sigma_e^2 \quad (8.43)$$

with the average variance without the covariate

$$\text{av. var}(\hat{\tau}_i - \hat{\tau}_{i'}) = \frac{2}{r} \sigma_e^2. \quad (8.44)$$

For (8.43) to be smaller than (8.44) we require

$$\rho^2 > 1 - \frac{t(r-1)}{t(r-1) + 1} \cdot \frac{t(r-1) - 2}{t(r-1) - 1}.$$

In Table 8.3 we give minimal values of ρ for selected values of t and r . These should be viewed as rough guidelines only, especially for small values of t and/or r , since there may be substantial variation in precision between different randomization patterns and between different comparisons within one randomization pattern (Cox, 1957).

The general conclusion from Table 8.3 is that for $|\rho| < .3$, the use of covariates is of no real value. Substantial gains will be realized, however, if $|\rho|$ is large.

8.5.2 Common Slope

Implicit in model (8.5) and Figure 8.2 is the assumption that the linear relationship between x and y is the same for all treatments, that is, the t regression lines are parallel. This is sometimes considered to be a serious and questionable assumption. This may be true for some situations in which the analysis of covariance is used, that is, observational studies, but should generally not be a problem in a CRD if proper randomization has taken place. The assumption is, however, checked easily using the procedure described below.

Consider the "full" model

$$y_{ij} = \mu_i + \beta_i(x_{ij} - \bar{x}_{..}) + \text{error}, \quad (8.45)$$

Table 8.3 Minimal Values for Correlation between Observations and Covariates in CRD

t	r	$ \rho $
2	5	.49
	10	.33
	15	.27
	20	.23
3	5	.40
	10	.27
	15	.22
4	5	.35
	10	.23
	15	.19
5	5	.35
	10	.21
6	5	.29
	10	.19
7	5	.27

where $\mu_i = \mu + \tau_i$ ($i = 1, 2, \dots, t; j = 1, 2, \dots, r$). We wish to test the hypothesis

$$H_0: \beta_1 = \beta_2 = \dots = \beta_t = \beta, \text{ say.}$$

We do this by fitting the model (8.45) and the “reduced” model (that is, assuming H_0 is true)

$$y_{ij} = \mu_i + \beta(x_{ij} - \bar{x}_{..}) + \text{error} \quad (8.46)$$

and obtain the sums of squares for both models, say SS_F and SS_R , respectively. We then test H_0 by considering the F -statistic

$$F = \frac{(SS_F - SS_R)/(t - 1)}{\left(\sum_{ij} y_{ij}^2 - SS_F \right) / t(r - 2)} \quad (8.47)$$

with $t - 1$ and $t(r - 2)$ d.f. The NE for model (8.45) are

$$r\hat{\mu}_i + r(\bar{x}_{i.} - \bar{x}_{..})\hat{\beta}_i = y_{i.} \quad (i = 1, 2, \dots, t)$$

$$r(\bar{x}_{i.} - \bar{x}_{..})\hat{\mu}_i + \sum_j (x_{ij} - \bar{x}_{..})^2 \hat{\beta}_i = \sum_j y_{ij}(x_{ij} - \bar{x}_{..}) \quad (i = 1, 2, \dots, t).$$

It then follows that

$$\begin{aligned}\hat{\mu}_i &= \bar{y}_i - \hat{\beta}_i(\bar{x}_i - \bar{x}_{..}) \\ \hat{\beta}_i &= \frac{\sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})}{\sum_j (x_{ij} - \bar{x}_{i.})^2} \equiv \frac{S_{ixy}}{S_{ixx}}, \text{ say}\end{aligned}\quad (8.48)$$

and

$$\begin{aligned}\text{SS}_F &= \sum_i \hat{\mu}_i \bar{y}_i + \sum_{ij} \hat{\beta}_i (x_{ij} - \bar{x}_{i.}) y_{ij} \\ &= r \sum_i \bar{y}_i^2 + \sum_i \frac{S_{ixy}^2}{S_{ixx}}\end{aligned}\quad (8.49)$$

with $2t$ d.f. Similarly, for model (8.46) the NE are

$$\begin{aligned}r\hat{\mu}_i + r(\bar{x}_i - \bar{x}_{..})\hat{\beta} &= y_i. \quad (i = 1, 2, \dots, t) \\ r \sum_i (\bar{x}_i - \bar{x}_{..})\hat{\mu}_i + \sum_{ij} (x_{ij} - \bar{x}_{i.})^2 \hat{\beta} &= \sum_{ij} y_{ij} (x_{ij} - \bar{x}_{i.}).\end{aligned}$$

It follows then [see also (8.16) and (8.15)] that

$$\begin{aligned}\hat{\mu}_i &= \bar{y}_i - \hat{\beta}(\bar{x}_i - \bar{x}_{..}) \\ \hat{\beta} &= \frac{\sum_i S_{ixy}}{\sum_i S_{ixx}} = \frac{E_{xy}}{E_{xx}}\end{aligned}\quad (8.50)$$

(we mention here in passing that it follows from (8.48) and (8.50) that the estimate of β is obtained by weighted pooling of the estimates of the individual β_i), and [see also (8.26)]

$$\text{SS}_R = r \sum_i \bar{y}_i^2 + \frac{\left(\sum_i S_{ixy} \right)^2}{\sum_i S_{ixx}} \quad (8.51)$$

with $t + 1$ d.f. The test statistic (8.47) then takes on the form

$$F = \frac{\frac{1}{\nu_1} \left[\sum_i \frac{S_{ixy}^2}{S_{ixx}} - \frac{\left(\sum_i S_{ixy} \right)^2}{\sum_i S_{ixx}} \right]}{\frac{1}{\nu_2} \left[\sum_i S_{iyy} - \sum_i \frac{S_{ixy}^2}{S_{ixx}} \right]} \quad (8.52)$$

with $\nu_1 = 2t - (t + 1) = t - 1$ and $\nu_2 = tr - 2t = t(r - 2)$ as in (8.47). If $F > F_{1-\alpha, t-1, t(r-2)}$ for a suitably chosen α , then we reject H_0 and conclude that the slopes are not all the same. Since this is considered to be a preliminary test we may choose $\alpha = .25$ rather than the customary $\alpha = .05$ (see Bancroft, 1964). For a numerical example of this procedure see Section 8.8.

If H_0 is rejected it might be useful to investigate the data more closely, for example by plotting or by a formal test to see if the nonparallelism is due to perhaps one treatment. One then may delete that treatment and proceed with the analysis of the other treatments in the usual fashion. If no such simple explanation is plausible it is difficult to prescribe what to do. In any case, model (8.5) is no longer appropriate, rather model (8.45) should then be used. In that case, however, treatment comparisons depend on the x -value at which they are compared and that may be rather unsatisfactory and misleading.

8.5.3 Covariates Affected by Treatments

To understand intuitively the problem that arises when the covariates are “affected” by the treatments, consider Figure 8.3.

In this case low y -values are associated with low x -values for T_1 (that is, treatment 1) and high y -values are associated with high x -values for T_2 (that is, treatment 2), the two treatments we may want to compare. As an example, suppose the treatments are varieties of potatoes and we want to compare the yield of these varieties using as a covariate the size of the seed potatoes. It so happens that variety 1 has small seed potatoes and variety 2 has large seed potatoes. If we were to apply the analysis of covariance procedure we would compare the varieties at seed potato size $x = \bar{x}$, a value which may not be achieved by either variety. Hence, this procedure is obviously of no value. Similar arguments apply to situations where the covariates are affected by the treatments in other ways. For an interesting discussion the reader is referred to Smith (1957).

We mentioned earlier that if the covariates are observed before the treatments are assigned they are certainly not affected by the treatments. But even in that case a situation as described in Figure 8.3 could arise for two reasons:

(i) due to a particular outcome of the randomization process and (ii) due to a lack of randomization. In situation (i) one should throw out the randomization pattern and repeat the randomization process; in situation (ii) one should expose the motives of the investigator. In practice, of course, one may not be able to distinguish between (i) and (ii). A method to protect oneself against this situation would be to subject the covariates x to an ANOVA for a CRD and consider $F = T_{xx}t(r-1)/[E_{xx}(t-1)]$ and if F is “large,” say larger than $F_{1-\alpha, t-1, t(r-1)}$ with $\alpha = .25$, assume that the covariates are “affected” by the treatments and proceed accordingly.

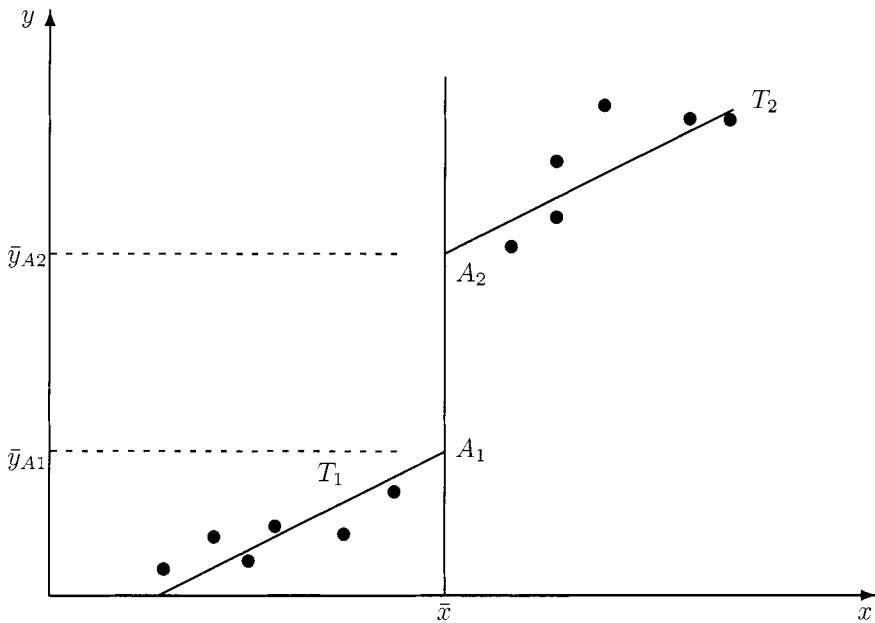


Figure 8.3 Covariates affected by treatments.

8.5.4 Normality Assumption

Even though we have not invoked the assumption of normality explicitly (we have instead argued that the tests based on normality are an approximation to randomization tests), we have used the method of least squares to estimate parameters of model (8.5). It is well known that the LS method is optimal only if the observations are normally distributed. Hence the question arises: What should one do if the observations are not normally distributed. The obvious answer is to replace the LS method by any of the available “robust” methods, e.g., M-estimation (Huber, 1964). A limited study along these lines was done by Birch and Myers (1982) for heavy-tailed error distributions. We quote from their conclusions:

The least squares procedures, although lacking efficiency in estimation of the parameters, show strong affinity toward the normal size in tests concerning parameter values. The t -like tests based on M-estimators can be studentized to follow the t -distribution. Due to the similarity of the test results for LS and M-estimators it is suggested that both procedures be used together to provide a basis of comparison and diagnostic examination of the data. If parameter estimates differ significantly then outliers in the data should be strongly suspected and may be examined. Tests on parameters can be performed based on least squares and/or M-estimates using the t -like or F -like procedures.

We add that the tests referred to are those given in Section 8.3.5 and that given in (8.52).

Along similar lines Hocking (1982) suggested to use regression diagnostics such as the so-called hat matrix and residuals to detect deviations from normality through e.g., high-leverage observations. We shall not pursue these arguments here but rather refer the reader to some of the relevant literature in this area, for example, Belsley, Kuh and Welsch (1980), Myers (1990). For nonparametric procedures we refer to Conover and Iman (1982).

8.6 ANALYSIS OF COVARIANCE WITH SUBSAMPLING

As an extension of the analysis of covariance procedure presented so far, we now discuss briefly the case involving subsampling (see Section 6.8). We can think of two situations:

- (i) The covariate is observable only for the EU; that is, x_{ij} .
- (ii) The covariate is observed for each OU, that is, x_{ijk} .

As an example of (i), consider a study to compare different drugs for their effectiveness in reducing blood pressure. Each patient (from a specified population) is given one of the drugs at random. To account for some variability among the patients, the blood pressure reading at the beginning of the trial is used as a covariate. At the end of the study duplicate blood pressure readings are obtained for each patient. An example for (ii) is the previously considered air pollution study (see Table 2.5) where each growth chamber represents the EU and the initial height of each plant (OU) in each chamber is used as a covariate. Remembering that the analysis of covariance is used as a device for reducing the experimental error it is only proper to treat both situations in the same way. For (i) we use the covariate x_{ij} as observed, and for (ii) we use as the covariate the average of the supplementary observations for each EU, that is, $x_{ij} = \bar{x}_{ij} = (1/n)\sum_k x_{ijk}$.

As an extension of (8.5), the model for such data can then be written as

$$y_{ijk} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}^* + \eta_{ijk} \quad (8.53)$$

where $i = 1, 2, \dots, t; j = 1, 2, \dots, r'; k = 1, 2, \dots, n$ and all the terms are as defined earlier (see Sections 8.3 and 6.8). Model (8.53) can be rewritten, for purposes of analysis, in terms of the average observation for the j th replication of treatment i as

$$\bar{y}_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + e_{ij}^{**} \quad (8.54)$$

where $e_{ij}^{**} = \varepsilon_{ij}^* + \bar{\eta}_{ij}$. The form of the model suggests that the basic analysis can be carried out as described in Section 8.3, substituting \bar{y}_{ij} for y_{ij} , $\sigma_{e^{**}}^2$ for σ_e^2 , and r' for r . More precisely, we obtain the entries in Table 8.4 using the \bar{y}_{ij} 's as the "observations." For example,

$$T_{yy} = r' \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$$

Table 8.4 ANOVA for Model (8.53)

Source	d.f.	SS
$\mathbf{X}_\tau \mathbf{J}, \mathbf{X}_\beta, \mathbf{X}_{\varepsilon^*}$	$t - 1$	$T_{yy}^* - \frac{S_{xy}^{*2}}{S_{xx}^*} + \frac{E_{xy}^{*2}}{E_{xx}^*}$
$\mathbf{X}_\beta \mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_{\varepsilon^*}$	1	$\frac{E_{xy}^{*2}}{E_{xx}^*}$
$\mathbf{X}_{\varepsilon^*} \mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta$	$t(r' - 1) - 1$	$E_{yy}^* - \frac{E_{xy}^{*2}}{E_{xx}^*}$
$\mathbf{I} \mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta, \mathbf{X}_{\varepsilon^*}$	$tr'(n - 1)$	O_{yy}^*
Total	$tr'n - 1$	

For purposes of the ANOVA table, however, we need to reconvert everything to a per-observation basis by simply defining

$$T_{yy}^* = nT_{yy}, \quad E_{yy}^* = nE_{yy}, \quad O_{yy}^* = \text{SS}(OE)$$

(see Table 6.8), and so on. The resulting ANOVA table is then as given in Table 8.4 using obvious notation. It should be clear, from our earlier discussion, how Table 8.4 can be used. For example, to test $H_0: \tau_1 = \tau_2 = \dots = \tau_t = 0$ we use

$$F = \frac{\text{MS}(\mathbf{X}_\tau | \mathbf{J}, \mathbf{X}_\beta, \mathbf{X}_{\varepsilon^*})}{\text{MS}(\mathbf{X}_{\varepsilon^*} | \mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta)}$$

with $t - 1$ and $t(r' - 1) - 1$ d.f. Furthermore, the sampling and experimental error variance components are estimated as

$$\hat{\sigma}_\eta^2 = \text{MS}(\mathbf{I} | \mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta, \mathbf{X}_{\varepsilon^*})$$

and

$$\hat{\sigma}_{\varepsilon^*}^2 = \frac{\text{MS}(\mathbf{X}_{\varepsilon^*} | \mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta) - \text{MS}(\mathbf{I} | \mathbf{J}, \mathbf{X}_\tau, \mathbf{X}_\beta, \mathbf{X}_{\varepsilon^*})}{n}.$$

8.7 CASE OF SEVERAL COVARIATES

In our discussion of the analysis of covariance technique so far we have considered the simplest but most important situation, namely that of one covariate and a linear relationship between the covariate x and the observation y . There may, however, be situations where the relationship between x and y is of a polynomial form or it may be useful to consider several covariates x_1, x_2, \dots which have a linear or polynomial

relationship with y . We have described and dealt with the general model involving both classificatory and regression parts in Section 4.14. Below we give a slightly different derivation following Cox and McCullagh (1982) (see also Scheffé, 1959).

8.7.1 General Case

Using matrix notation we write the general analysis of covariance model for the $N \times 1$ vector of observations \mathbf{y} as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}^*, \quad (8.55)$$

where $\mathbf{X}\boldsymbol{\mu}$ represents the classificatory part (treatments in our case) and $\mathbf{Z}\boldsymbol{\gamma}$ represents the regression part (the covariates in our case), \mathbf{X} and \mathbf{Z} are matrices of known constants of dimensions $N \times d_x$ and $N \times d_z$, respectively, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are $d_x \times 1$ and $d_z \times 1$ vectors, respectively, of unknown parameters, and \mathbf{e}^* is a $N \times 1$ vector of errors with $E(\mathbf{e}^*) = \mathbf{0}$, and $\text{var}(\mathbf{e}^*) = \mathbf{I}\sigma_{e^*}^2$. If no covariates are included or, alternatively, if $\boldsymbol{\gamma} = \mathbf{0}$, then model (8.55) reduces to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{e}. \quad (8.56)$$

We shall refer to (8.56) also as the *design model*. We know (see Chapter 4) that for (8.56) an orthogonal decomposition of \mathbf{y} is given by

$$\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} = \mathbf{P}_\mathbf{X}\mathbf{y} + \mathbf{R}_\mathbf{X}\mathbf{y}, \quad (8.57)$$

where

$$\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad \text{and} \quad \mathbf{R}_\mathbf{X} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \mathbf{I} - \mathbf{P}_\mathbf{X}$$

are $N \times N$ idempotent matrices (we assume here that the parameterization in (8.55) and (8.56) is such that $\text{rank}(\mathbf{X}) = d_x$). In (8.57) $\mathbf{R}_\mathbf{X}\mathbf{y}$ is the vector of residuals and $\mathbf{y}'\mathbf{R}_\mathbf{X}\mathbf{y}$ is the residual sum of squares. We now rewrite (8.55) as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\mu}_{(0)} + \mathbf{R}_\mathbf{X}\mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}^* \\ &= \mathbf{X}[\boldsymbol{\mu}_{(0)} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\gamma}] + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}^* \end{aligned} \quad (8.58)$$

so that

$$\boldsymbol{\mu} = \boldsymbol{\mu}_{(0)} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\gamma}. \quad (8.59)$$

Using (8.58) the NE are obtained as

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{R}_\mathbf{X}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}_\mathbf{X}\mathbf{X} & \mathbf{Z}'\mathbf{R}_\mathbf{X}\mathbf{Z} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\mu}}_{(0)} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{R}_\mathbf{X}\mathbf{y} \end{bmatrix},$$

which reduces to

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \emptyset \\ \emptyset & \mathbf{Z}'\mathbf{R}_\mathbf{X}\mathbf{Z} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\mu}}_{(0)} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{R}_\mathbf{X}\mathbf{y} \end{bmatrix}. \quad (8.60)$$

From (8.60) we obtain immediately

$$\hat{\boldsymbol{\mu}}_{(0)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (8.61)$$

that is, the estimator for μ under the design model,

$$\hat{\gamma} = (\mathbf{Z}'\mathbf{R}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}_X\mathbf{y} \quad (8.62)$$

and from (8.59)

$$\hat{\mu} = \hat{\mu}_{(0)} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\hat{\gamma}. \quad (8.63)$$

It also follows from the special form of (8.60) that

$$\text{var}(\hat{\mu}_{(0)}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma_{e^*}^2$$

and

$$\text{var}(\hat{\gamma}) = (\mathbf{Z}'\mathbf{R}_X\mathbf{Z})^{-1}\sigma_e^2. \quad (8.64)$$

and, since $\hat{\mu}_{(0)}$ and $\hat{\gamma}$ are uncorrelated,

$$\text{var}(\hat{\mu}) = [(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{Z}'\mathbf{R}_X\mathbf{Z})^{-1}]\sigma_{e^*}^2. \quad (8.65)$$

We comment briefly on the form of (8.62). The elements of $\mathbf{Z}'\mathbf{R}_X\mathbf{Z}$ are the error sums of squares (diagonal elements) and error sums of products (off-diagonal elements) for the design model (8.56) when the columns of \mathbf{Z} are used as the “observation” vectors. Similarly, the elements of the vector $\mathbf{Z}'\mathbf{R}_X\mathbf{y}$ are the corresponding error sums of products using successively the columns of \mathbf{Z} with the observation vector \mathbf{y} . This presents an easy way of obtaining $\hat{\gamma}$ and hence $\hat{\mu}$ as we shall illustrate in Section 8.7.2.

The error sum of squares, $SS(\mathbf{I}|\mathbf{X}, \mathbf{Z})$, is obtained in the usual way as

$$SS(\mathbf{I}|\mathbf{X}, \mathbf{Z}) = \mathbf{y}'\mathbf{y} - \hat{\mu}'_{(0)}\mathbf{X}'\mathbf{y} - \hat{\gamma}'\mathbf{Z}'\mathbf{R}_X\mathbf{y}. \quad (8.66)$$

It is instructive to write (8.66) as

$$SS(\mathbf{I}|\mathbf{X}, \mathbf{Z}) = SS(\mathbf{I}|\mathbf{X}) - \hat{\gamma}'\mathbf{Z}'\mathbf{R}_X\mathbf{y}$$

which shows that the error sum of squares for model (8.55) is smaller than the error sum of squares for model (8.56), and the reduction is given by $\hat{\gamma}'\mathbf{Z}'\mathbf{R}_X\mathbf{y}$. From (8.66) we then obtain

$$\begin{aligned} \hat{\sigma}_{e^*}^2 &= MS(\mathbf{I}|\mathbf{X}, \mathbf{Z}) \\ &= SS(\mathbf{I}|\mathbf{X}, \mathbf{Z}) / (N - d_x - d_z). \end{aligned}$$

Finally, to test any hypothesis about μ or a subvector of

$$\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix},$$

say $H_0: \mu_{(1)} = \mu^*$, we fit the model

$$\mathbf{y} = \mathbf{X}^* \begin{pmatrix} \mu^* \\ \mu_{(2)} \end{pmatrix} + \mathbf{Z}\gamma + \mathbf{e}^* \quad (8.67)$$

say, and obtain $SS(\mathbf{I}|\mathbf{X}^*, \mathbf{Z})$ in the same way as we obtained (8.66). Suppose H_0 is of rank d . We then form the F -statistic

$$F = \frac{[SS(\mathbf{I}|\mathbf{X}^*, \mathbf{Z}) - SS(\mathbf{I}|\mathbf{X}, \mathbf{Z})]/d}{MS(\mathbf{I}|\mathbf{X}, \mathbf{Z})} \quad (8.68)$$

or alternatively,

$$F = \frac{[SS(\mathbf{X}, \mathbf{Z}) - SS(\mathbf{X}^*, \mathbf{Z})]/d}{MS(\mathbf{I}|\mathbf{X}, \mathbf{Z})}$$

with d and $N - d_x - d_z$ d.f. For $\mu_1^* = 0$ this procedure is derived explicitly in Section 4.14.2.

8.7.2 Two Covariates

We shall illustrate the procedure described above in terms of a simple example in the context of the CRD. Suppose we have t treatments, each replicated r times, and two covariates x and z (for polynomial regression with one covariate x and a quadratic relationship, we can use this technique by taking the second “covariate” to be $z = x^2$). Then, in model (8.55) we have

$$\boldsymbol{\mu}' = (\mu_1, \mu_2, \dots, \mu_t),$$

where for the CRD we have $\mu_i = \mu + \tau_i (i = 1, 2, \dots, t)$,

$$\mathbf{X} = \begin{bmatrix} \mathbf{J}_r & & & \\ & \mathbf{J}_r & & \\ & & \ddots & \\ & & & \mathbf{J}_r \end{bmatrix}$$

where \mathbf{J}_r is a $r \times 1$ column vector of unity elements and \mathbf{X} contains t such vectors,

$$\boldsymbol{\gamma}' = (\gamma_1, \gamma_2)$$

$$\mathbf{Z} = \begin{bmatrix} x_{11}^* & z_{11}^* \\ x_{12}^* & z_{12}^* \\ \vdots & \vdots \\ x_{tr}^* & z_{tr}^* \end{bmatrix},$$

where $x_{ij}^* = x_{ij} - \bar{x}_{..}$, $z_{ij}^* = z_{ij} - \bar{z}_{..}$, and x_{ij} and z_{ij} are the covariates for the j th replication of treatment i . In a practical setting the treatments may be different advertising strategies for a book of general interest, the EUs are comparable book stores in different cities, x may be the sales volume of a bookstore in the previous month, z may be the price of the book established prior to the advertising campaign (there being slight differences in price due to local conditions), and y being the sales volume of this book during a specified period.

We then find from (8.61)

$$\hat{\boldsymbol{\mu}}_{(0)} = (\bar{y}_{1.}, \bar{y}_{2.}, \dots, \bar{y}_{t.})',$$

and from (8.62)

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} E_{xx} & E_{xz} \\ E_{zx} & E_{zz} \end{pmatrix}^{-1} \begin{pmatrix} E_{xy} \\ E_{zy} \end{pmatrix},$$

where E_{xz}, E_{zx}, E_{zy} are defined in an obvious way as an extension of the terms in Table 8.1 and as described above as error sums of squares and error sums of products for the CRD. Further, from (8.63) we obtain the i th component of $\hat{\mu}$ as

$$\begin{aligned} \hat{\mu}_i &= \bar{y}_i - \hat{\gamma}_1 \bar{x}_i^* - \hat{\gamma}_2 \bar{z}_i^* \\ &= \bar{y}_i - \hat{\gamma}_1 (\bar{x}_i - \bar{x}_{..}) - \hat{\gamma}_2 (\bar{z}_i - \bar{z}_{..}). \end{aligned}$$

Also, from (8.66) we find

$$\begin{aligned} \text{SS(Error)} &= \text{SS}(\mathbf{I}|\mathbf{X}, \mathbf{Z}) \\ &= E_{yy} - E_{xy}\hat{\gamma}_1 - E_{zy}\hat{\gamma}_2 \end{aligned}$$

and

$$\hat{\sigma}_{e^*}^2 = (E_{yy} - E_{xy}\hat{\gamma}_1 - E_{zy}\hat{\gamma}_2) / [t(r-1) - 2]. \quad (8.69)$$

Finally, to test the equality of treatment effects $H_0: \mu_1 = \mu_2 = \dots = \mu_t = \mu$, model (8.68) takes on the form

$$\mathbf{y} = \mathbf{J}\mu + \mathbf{Z}\gamma + \mathbf{e}^*.$$

The NE then yield the estimators $\tilde{\mu}, \tilde{\gamma}_1, \tilde{\gamma}_2$ for μ, γ_1, γ_2 , respectively, as

$$\begin{aligned} \tilde{\mu} &= \bar{y}_{..} \\ \begin{pmatrix} S_{xx} & S_{xz} \\ S_{zx} & S_{zz} \end{pmatrix} \begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} &= \begin{pmatrix} S_{xy} \\ S_{zy} \end{pmatrix} \end{aligned}$$

and hence

$$\text{SS}(\mathbf{I}|\mathbf{J}, \mathbf{Z}) = S_{yy} - S_{xy}\tilde{\gamma}_1 - S_{zy}\tilde{\gamma}_2$$

so that (8.68) becomes

$$F = \frac{(T_{yy} - S_{xy}\tilde{\gamma}_1 - S_{zy}\tilde{\gamma}_2 + E_{xy}\hat{\gamma}_1 + E_{zy}\hat{\gamma}_2) / (t-1)}{(E_{yy} - E_{xy}\hat{\gamma}_1 - E_{zy}\hat{\gamma}_2) / [t(r-1) - 2]}.$$

Tests for γ_1 and γ_2 can, of course, be derived in a similar fashion.

The general case with m covariates, x_1, x_2, \dots, x_m say, is discussed in Section 4.14.4. It is shown there, as is somewhat intuitive from the case $m = 2$ above, that the arithmetic can be represented by the ANOVA of $\mathbf{y}, \mathbf{X}_1, \dots, \mathbf{X}_m$ and the corresponding sums of products.

Many of the problems discussed in Section 8.5 can arise in the multiple covariate situation as well and extra care must be taken to assure validity of the basic assumptions. For example, problems of collinearity may arise and appropriate diagnostics and/or different estimators for the regression coefficients in γ may have to be used as described for example by Myers (1990). The problem may become rather complicated and in the end not worth the effort as there may be only marginal reduction of error as the number of covariates increases.

8.8 EXAMPLES USING SAS®

Even though the analysis of covariance in its simplest form is easy to perform, computer programs have, nevertheless, led to a much wider and more common use of this method of reducing experimental error. In the following we shall give some examples and illustrate the use of SAS Proc GLM.

EXAMPLE 8.1: Consider the experimental situation described in Exercise 8.3 with the data given also in Table 8.5a. The input statements for SAS PROC GLM are given in Table 8.5a. There we have included several options such as “inverse”, “solution”, and “e” the reason for which we shall explain in the comments for the output, given in Table 8.5b:

- (i) The Type III SS in the ANOVA show that there are differences among the treatments ($P < .0001$) and that the regression coefficient is different from zero ($P = 0.0005$).
- (ii) The solution vector (which is produced because of the “solution” option, required for classificatory models) gives $\hat{\beta} = .773$ with standard error $se(\hat{\beta}) = 0.16$.
- (iii) The $se(\hat{\beta})$ can also be obtained from the $\mathbf{X}'\mathbf{X}$ Generalized Inverse as

$$se(\hat{\beta}) = (0.03358 \times 0.7656)^{1/2}$$

where $.7656 = \hat{\sigma}_{e*}^2$ from the ANOVA table.

- (iv) The General Form of Estimable Functions (obtained because of the option “e” in the model statement) can be used to interpret the solutions for the treatment effects. For example, for treatment 1 we have

$$6.2245 = \widehat{\tau_1 - \tau_3},$$

which is obtained by putting $L2 = 1$ and all other $Li = 0$, with

$$se(\widehat{\tau_1 - \tau_3}) = 0.6021.$$

- (v) The generalized inverse can also be used to obtain, for example,

$$se(\widehat{\tau_1 - \tau_3}) = [(.4735 + .4113 - 2 \times .1711) \times .7656]^{1/2}.$$

- (vi) The “e” option for LSmeans shows us how to obtain the LSmeans. To do so, however, we need to mention that instead of model (8.5) SAS uses the model

$$y_{ij} = \mu^* + \tau_i + \beta x_{ij} + e_{ij}^*.$$

Table 8.5 Basic Analysis of Covariance

a) Input statements:

```
data ancova;
input trt x y @@;
datalines;
1 1 57.0 1 2 55.0 1 3 62.1 1 4 74.5 1 5 86.7 1 6 42.0
2 1 64.8 2 2 66.6 2 3 69.5 2 4 61.1 2 5 91.8 2 6 51.8
3 1 70.7 3 2 59.4 3 3 64.5 3 4 74.0 3 5 78.5 3 6 55.8
4 1 68.3 4 2 67.1 4 3 69.1 4 4 72.7 4 5 90.6 4 6 44.3
5 1 76.0 5 2 74.5 5 3 76.5 5 4 86.6 5 5 94.7 5 6 43.2
;
run;

proc print data=ancova;
title 'DATA FOR CRD WITH SUPPLEMENTARY INFORMATION';
run;

proc glm data=ancova;
class trt;
model y=trt x/ inverse solution e;
means trt;
lsmeans trt/stderr e;
title 'BASIC ANALYSIS OF COVARIANCE';
run;
```

b) Output:

DATA FOR CRD WITH SUPPLEMENTARY INFORMATION			
Obs	trt	x	y
1	1	4.1	12.5
2	1	2.9	10.3
3	1	1.5	9.6
4	1	4.3	12.6
5	1	2.2	11.3
6	2	6.8	11.5
7	2	2.7	8.6
8	2	3.8	7.2
9	2	6.4	11.6
10	2	5.6	8.9
11	3	6.6	6.8
12	3	2.2	4.8
13	3	3.5	5.6
14	3	5.5	7.5
15	3	4.6	6.2

Table 8.5 (Continued)

BASIC ANALYSIS OF COVARIANCE				
The GLM Procedure				
Class Level Information				
Class	Levels	Values		
trt	3	1	2	3
Number of Observations Read				
Number of Observations Used				
X'X Generalized Inverse (g2)				
	Intercept	trt 1	trt 2	
Intercept	0.8739556749	-0.422646071	-0.11274681	
trt 1	-0.422646071	0.4735527199	0.1711752854	
trt 2	-0.11274681	0.1711752854	0.4112961719	
trt 3	0	0	0	
x	-0.150436535	0.0496977837	-0.019476158	
y	2.7154466085	6.2245399597	2.9314640698	
X'X Generalized Inverse (g2)				
	trt 3	x	y	
Intercept	0	-0.150436535	2.7154466085	
trt 1	0	0.0496977837	6.2245399597	
trt 2	0	-0.019476158	2.9314640698	
trt 3	0	0	0	
x	0	0.0335795836	0.7733378106	
y	0	0.7733378106	8.4220302216	
General Form of Estimable Functions				
Effect	Coefficients			
Intercept	L1			
trt	1	L2		
trt	2	L3		
trt	3	L1-L2-L3		
x	L5			

Table 8.5 (Continued)

BASIC ANALYSIS OF COVARIANCE					
The GLM Procedure					
Dependent Variable: y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	84.67796978	28.22598993	36.87	<.0001
Error	11	8.42203022	0.76563911		
Corrected Total	14	93.10000000			
	R-Square	Coeff Var	Root MSE	y Mean	
	0.909538	9.722312	0.875008	9.000000	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	2	66.86800000	33.43400000	43.67	<.0001
x	1	17.80996978	17.80996978	23.26	0.0005
Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	83.14658219	41.57329109	54.30	<.0001
x	1	17.80996978	17.80996978	23.26	0.0005
Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		2.715446608 B	0.81800651	3.32	0.0068
trt	1	6.224539960 B	0.60213826	10.34	<.0001
trt	2	2.931464070 B	0.56116347	5.22	0.0003
trt	3	0.000000000 B	.	.	.
x		0.773337811	0.16034289	4.82	0.0005

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Table 8.5 (Continued)

The GLM Procedure					
Level of trt	N	-----y-----		-----x-----	
		Mean	Std Dev	Mean	Std Dev
1	5	11.2600000	1.32400906	3.00000000	1.20415946
2	5	9.5600000	1.92691463	5.06000000	1.75157072
3	5	6.1800000	1.04498804	4.48000000	1.71084774
Least Squares Means					
Coefficients for trt Least Square Means					
Effect		trt Level			
		1	2	3	
Intercept		1	1	1	
trt	1	1	0	0	
trt	2	0	1	0	
trt	3	0	0	1	
x		4.18	4.18	4.18	
		trt	y LSMEAN	Standard Error	Pr > t
		1	12.1725386	0.4346564	<.0001
		2	8.8794627	0.4159778	<.0001
		3	5.9479987	0.3942610	<.0001

Then the coefficients for trt Least Square Means tells us that, for example,

$$\begin{aligned} \text{LSmean(trt 1)} &= \tilde{\mu} + \tilde{\tau}_1 + \hat{\beta} \cdot 4.18 \\ &= 2.7154 + 6.2245 + .7733 \cdot 4.18 \\ &= 12.1723, \end{aligned}$$

which is, apart from rounding error, equal to the LSmean (trt 1) given in the SAS output. Here $\tilde{\mu}$ and $\tilde{\tau}_1$ are part of the solutions to the NE obtained by SAS (using $\tilde{\tau}_3 = 0$), and $4.18 = \bar{x}_\cdot$.

- (vii) The $\text{se}[\text{LS mean (trt 1)}]$ can be obtained by using the generalized inverse, $\hat{\sigma}_e^2$ -, and the coefficients for the LS mean as

$$\begin{aligned} \text{se}[\text{LSmean(trt) 1}] &= [(.8740 + .4736 + (4.18)^2 \times \\ &\quad .0336 - 2 \times .4226 - 2 \times .1504 \\ &\quad + 2 \times 4.18 \times .0497) \times .7656]^{1/2} = .435 \end{aligned}$$

- (viii) Just for comparison we give y -means in addition to the LS means to show that LS mean for trt 1 is adjusted upwards, whereas those for trt 2 and 3 are adjusted downwards as an illustration of Figure 8.2.

EXAMPLE 8.2: We consider the same data as in Example 8.1. In addition to the analysis of covariance we now consider post-hoc comparisons in the form of orthogonal contrast and illustrate the Tukey-Kramer procedure for multiple comparisons (see Section 8.4.2). The input statements are given in Table 8.6a and the output in Table 8.6b.

We make the following comments:

- (i) In the input statement, in order to perform the Tukey procedure for the LSmeans we have to specify “adjust=Tukey” (other multiple comparison procedures are available, but not Duncan’s multiple range test).
- (ii) The Tukey-Kramer adjustments apply, of course, only to the multiple comparison tests and simultaneous confidence intervals.
- (iii) The low P -values, indicating highly significant differences between the treatments, correspond to lower and upper confidence limits having the same sign for each pairwise comparison.
- (iv) Note that the contrast sums of squares do not add up to the treatment sum of squares (see Section 7.3). \square

EXAMPLE 8.3: Using the data from Example 8.1 we shall demonstrate the use of SAS PROC GLM to obtain separate regression coefficients for each treatment and then test for equality of slopes. This is done in two steps with the input statements given in Table 8.7a.

For fitting separate regression lines we consider the regressions, technically speaking, to be nested within treatments, expressed as “ $x(\text{trt})$ ”. This is equivalent to model (8.45). The procedure for testing for equality of slopes is different than the method described in Section 8.5.2. The input statement “ $x\ x*\text{trt}$ ” results in “fitting” first a single slope, indicated by “ x ”, and then considers deviations from the single slope, indicated by “ $x*\text{trt}$ ”. Testing $x*\text{trt} = 0$ is then equivalent to testing $H_0 : \beta_1 = \beta_2 = \dots = \beta_t$.

The results of these two procedures are given in Table 8.7b:

- (i) Using the “solution” option provides $\hat{\beta}_1 = .976$, $\hat{\beta}_2 = .896$, $\hat{\beta}_3 = .545$ as the estimates of the three regression coefficients.
- (ii) The results in (i) may suggest that the regression coefficients may not be equal, but the test for “ $x*\text{trt}$ ” is not significant ($P = .5548$), from which we conclude that the assumption of a common slope for the three treatments is reasonable.
- (iii) Looking at the solution vector for x and $x*\text{trt}$ we recognize that the single slope mentioned above is actually $\hat{\beta}_3 = .5448$ and $x*\text{trt } 1 = \hat{\beta}_1 - \hat{\beta}_3 = .4311$; $x*\text{trt}2 = \hat{\beta}_2 - \hat{\beta}_3 = .3509$. \square

Table 8.6 Analysis of Covariance with Post-Hoc Comparisons

a) Input statements:

```
data ancova;
input trt x y @@;
datalines;
1 4.1 12.5 1 2.9 10.3 1 1.5 9.6 1 4.3 12.6 1 2.2 11.3
2 6.8 11.5 2 2.7 8.6 2 3.8 7.2 2 6.4 11.6 2 5.6 8.9
3 6.6 6.8 3 2.2 4.8 3 3.5 5.6 3 5.5 7.5 3 4.6 6.2
;
run;

proc glm data=ancova;
class trt;
model y=trt x;
lsmeans trt/stderr pdiff cl adjust=Tukey ;
contrast '1+2 vs 3' trt 1 1 -2;
estimate '1+2 vs 3' trt 1 1 -2/divisor=2;
contrast '1 vs 2' trt 1 -1;
estimate '1 vs 2' trt 1 -1;
title1 'ANALYSIS OF COVARIANCE';
title2 'WITH POST-HOC COMPARISONS';
run;
```

b.) Output:

ANALYSIS OF COVARIANCE WITH POST-HOC COMPARISONS					
The GLM Procedure					
Class Level Information					
Class	Levels	Values			
trt	3	1 2 3			
Number of Observations Read				15	
Number of Observations Used				15	
Dependent Variable: y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	84.67796978	28.22598993	36.87	<.0001
Error	11	8.42203022	0.76563911		
Corrected Total	14	93.10000000			
Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	2	66.86800000	33.43400000	43.67	<.0001
x	1	17.80996978	17.80996978	23.26	0.0005
Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	83.14658219	41.57329109	54.30	<.0001
x	1	17.80996978	17.80996978	23.26	0.0005

Table 8.6 (Continued)

Least Squares Means Adjustment for Multiple Comparisons: Tukey-Kramer					
trt	y LSMEAN	Standard Error	Pr > t	LSMEAN Number	
1	12.1725386	0.4346564	<.0001	1	
2	8.8794627	0.4159778	<.0001	2	
3	5.9479987	0.3942610	<.0001	3	
Least Squares Means for effect trt Pr > t for H0: LSMean(i)=LSMean(j)					
Dependent Variable: y					
i/j	1	2	3		
1		0.0009	<.0001		
2	0.0009		0.0008		
3	<.0001	0.0008			
trt	y LSMEAN	95% Confidence Limits			
1	12.172539	11.215866	13.129211		
2	8.879463	7.963902	9.795024		
3	5.947999	5.080236	6.815761		
Least Squares Means for Effect trt					
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)		
1	2	3.293076	1.552453	5.033699	
1	3	6.224540	4.598281	7.850799	
2	3	2.931464	1.415870	4.447058	
ANALYSIS OF COVARIANCE WITH POST-HOC COMPARISONS					
The GLM Procedure					
Dependent Variable: y					
Contrast	DF	Contrast SS	Mean Square	F Value Pr > F	
1+2 vs 3	1	68.31196748	68.31196748	89.22	<.0001
1 vs 2	1	19.98964494	19.98964494	26.11	0.0003
Parameter	Estimate	Standard Error	t Value	Pr > t	
1+2 vs 3	4.57800201	0.48466275	9.45	<.0001	
1 vs 2	3.29307589	0.64448269	5.11	0.0003	

Table 8.7 Analysis of Covariance: Fitting Separate Regressions and Testing for Equality of Slopes

a) Input statements:

```
data ancova;
input trt x y @@;
datalines;
1 4.1 12.5 1 2.9 10.3 1 1.5 9.6 1 4.3 12.6 1 2.2 11.3
2 6.8 11.5 2 2.7 8.6 2 3.8 7.2 2 6.4 11.6 2 5.6 8.9
3 6.6 6.8 3 2.2 4.8 3 3.5 5.6 3 5.5 7.5 3 4.6 6.2
;
run;

proc glm data=ancova;
class trt;
model y=trt x(trt)/solution;
title1 'CRD WITH SUPPLEMENTARY INFORMATION';
title2 'FITTING SEPARATE REGRESSION LINES';
run;

proc glm data=ancova;
class trt;
model y=trt x x*trt/solution;
title2 'TESTING FOR EQUALITY OF SLOPES';
run;
```

b.) Output:

CRD WITH SUPPLEMENTARY INFORMATION					
FITTING SEPARATE REGRESSION LINES					
The GLM Procedure					
Dependent Variable: y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	85.71133848	17.14226770	20.88	0.0001
Error	9	7.38866152	0.82096239		
Corrected Total	14	93.10000000			
R-Square		Coeff Var	Root MSE	y Mean	
0.920637		10.06744	0.906070	9.000000	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	2	66.86800000	33.43400000	40.73	<.0001
x(trt)	3	18.84333848	6.28111283	7.65	0.0076
Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	6.13915023	3.06957511	3.74	0.0658
x(trt)	3	18.84333848	6.28111283	7.65	0.0076

Table 8.7 (Continued)

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		3.739494363 B	1.25360461	2.98	0.0154
trt	1	4.592919430 B	1.73482684	2.65	0.0266
trt	2	1.288276172 B	1.85702075	0.69	0.5054
trt	3	0.000000000 B	.	.	.
x(trt)	1	0.975862069	0.37622499	2.59	0.0290
x(trt)	2	0.895697523	0.25864492	3.46	0.0071
x(trt)	3	0.544755723	0.26480140	2.06	0.0698

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

CRD WITH SUPPLEMENTARY INFORMATION
TESTING FOR EQUALITY OF SLOPES

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	85.71133848	17.14226770	20.88	0.0001
Error	9	7.38866152	0.82096239		
Corrected Total	14	93.10000000			

R-Square	Coeff Var	Root MSE	y Mean
0.920637	10.06744	0.906070	9.000000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	2	66.86800000	33.43400000	40.73	<.0001
x	1	17.80996978	17.80996978	21.69	0.0012
x*trt	2	1.03336870	0.51668435	0.63	0.5548

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	6.13915023	3.06957511	3.74	0.0658
x	1	17.20712309	17.20712309	20.96	0.0013
x*trt	2	1.03336870	0.51668435	0.63	0.5548

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		3.739494363 B	1.25360461	2.98	0.0154
trt	1	4.592919430 B	1.73482684	2.65	0.0266
trt	2	1.288276172 B	1.85702075	0.69	0.5054
trt	3	0.000000000 B	.	.	.
x		0.544755723 B	0.26480140	2.06	0.0698
x*trt	1	0.431106346 B	0.46007067	0.94	0.3732
x*trt	2	0.350941800 B	0.37015804	0.95	0.3678
x*trt	3	0.000000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

8.9 EXERCISES

8.1 Consider the following data (y, x) from a CRD, where y represents the response after treatment, and x is a covariate.

Treatment	1:	(7.5, 3.5) (6.2, 2.6) (6.8, 3.1)
	2:	(10.0, 3.0) (12.1, 3.7) (11.3, 4.1)
	3:	(15.2, 5.1) (10.7, 2.6) (12.9, 3.1)
	4:	(5.1, 4.2) (4.6, 3.7) (7.1, 4.9)
	5:	(12.1, 2.3) (14.2, 2.9) (15.0, 3.5)
	6:	(10.0, 3.2) (9.8, 3.0) (9.6, 2.5)

(i) Analyze the data, ignoring the covariate, that is,

- (a) obtain treatment means,
- (b) obtain ANOVA table and perform F -test,
- (c) perform Tukey's Test ($\alpha = .05$),
- (d) interpret the results.

(ii) Using the experiment as a pilot study and, again ignoring the covariate, determine the number of replications per treatment needed to detect a difference between the best and the poorest treatment of 3 units or more with probability .8, using a test of size $\alpha = .05$.

(iii) Do the same as in (i) using the covariate.

(iv) Do the same as in (ii) using the covariate.

(v) Comment on the results from (i), (ii) vs. (iii), (iv).

8.2 An experiment was conducted to compare six different management techniques (such as pruning, spraying and fertilizing) for apple trees with respect to yield. Each apple tree represents an experimental unit and the trial was layed out as a completely randomized design with 5 replications for each management technique. All the trees underwent the same management practice before the trial. For each tree the yield in bushels (x) for the four-year period preceding the trial is available. At the end of the four-year experimental period, the yield in pounds of apples (y) is obtained for every tree.

Suppose the partial SAS Proc GLM printout is as follows:

Source	Type I SS	Type III SS
Treatments	40	60
Prev. Yield	26	26
Error	46	
Total	112	

Based on this information indicate how you would answer the following questions:

- (i) Are there differences among the management techniques (treatments)?
- (ii) Has the use of a covariate been successful in reducing the variance for treatment comparisons?
- (iii) Suppose the supplementary information were not available. How would you test $H_0: \tau_1 = \tau_2 = \dots = \tau_6$?
- (iv) Suppose $\bar{x}_{1.} = 10, \bar{x}_{2.} = 12, \bar{x}_{3.} = 9, \bar{x}_{4.} = 10, \bar{x}_{5.} = 13, \bar{x}_{6.} = 8, \Sigma(x_{ij} - \bar{x}_{i.})^2 = 20$. What is the standard error of the comparison “treatment 1 vs. remaining treatments?”

8.3 Suppose an engineer is interested in comparing three chemical processes for manufacturing a certain compound. She suspects that the impurity of the raw material used in the processes will affect the final product. She therefore wants to adjust for that in the final analysis.

Using a CRD with 15 experimental units she records the following:

Treatment	Amount of Impurity	Yield
1	4.1	12.5
	2.9	10.3
	1.5	9.6
	4.1	12.6
	2.2	11.3
2	6.8	11.5
	2.7	8.6
	3.8	7.2
	6.4	11.6
	5.6	8.9
3	6.6	6.8
	2.2	4.8
	3.5	5.6
	5.5	7.5
	4.6	6.2

- (i) Plot the data.
Using the methods and formulae described in Section 8.5.2,
- (ii) Estimate the regression line for each treatment.
- (iii) Test the hypotheses that the three slopes in (ii) are equal.
- (iv) Obtain the pooled estimate of the slope.
- (v) Obtain the unadjusted and the adjusted treatment means and compare them.
- (vi) Obtain the ANOVA table.
- (vii) Interpret the results obtained from the ANOVA table.
Compare the results with those obtained in Examples 8.1 – 8.3.

8.4 Show that

$$\text{av. var}(\hat{\tau}_i - \hat{\tau}_{i'}) = \frac{2}{r} \left[1 + \frac{1}{t-1} \frac{T_{xx}}{E_{xx}} \right] \sigma_{e^*}^2$$

[see (8.39)].

8.5 Show explicitly that $\hat{\beta}$ as given in (8.50) is the weighted average of the estimates of the individual $\hat{\beta}_i$ where the weights are the reciprocals of $\text{var}(\hat{\beta}_i)$ ($i = 1, 2, \dots, t$).

CHAPTER 9

Randomized Block Designs

9.1 INTRODUCTION

As we have mentioned on several occasions, one of the major objectives of considering designed experimentation is to reduce error in order to improve the sensitivity or precision of the investigation. Hence comes our use of the word error-reduction or error-control design (see Chapter 2) for an experimental plan which produces an error variance smaller than that for a comparable CRD for purposes of treatment comparisons. In Chapter 8 we have already considered one method of reducing the error. This was not achieved through a more complex design but rather by making effective use of additional information.

In this chapter we shall consider the situation referred to in Section 8.1 where the variability among the experimental units available for a study is systematic rather than “random.” Such variation may arise naturally or may be “induced” or introduced by the experimenter. Both situations are treated identically from the design point of view but may have to be treated differently from an analysis point of view. We shall discuss this later but shall give first some examples of both situations.

In a field experiment there may be a fertility gradient (due to sloping land, or example) such that EUs on the same gradient level are more alike than those at different levels; or there may be a creek running through the field such that plots equidistant from the creek are more alike than those at different distances from the creek (Pearce, 1983). In a clinical trial, to achieve adequate numbers of replications, several centers may be involved and patients (EUs) in the same center may be more alike than patients from different centers, not so much because of their own personal characteristics, but because of different treatment practices or management styles in different centers.

Induced variability is often considered when one wants to broaden the scope of the validity of experimental findings. An investigator in an industrial experiment may decide to obtain experimental material from different suppliers who use different production processes. In a livestock feeding trial it may be important to include animals from different breeds; or in an experiment to test different brands of tires one may want to include cars from different manufacturers and different models for each manufacturer.

It should be clear from these examples that there are many situations with systematic variation among sets of EUs. In such situations it is obviously inappropriate to use a CRD and the aim of designing the experiment must be to take this variation into account and to “eliminate” the effect such variability would have on the precision of treatment comparisons. This leads us to the concept of *local control* or *blocking* alluded to in Chapter 2. This concept, introduced by R. A. Fisher (1926, 1935), is indeed one of the most important concepts in the subject of experimental design and all error-control designs discussed in this chapter and the following chapters make use of it in one form or another.

To conclude this section we shall relate our discussion above to our general development in Section 2.2.4, where we divided the set of blocking factors in intrinsic factors, denoted by \mathcal{Z} , and non-specific factors, denoted by \mathcal{U} . An example of an intrinsic factor is given by the inclusion of different breeds in a feeding trial, whereas an example of a non-specific factor is the recognition of a fertility gradient due to sloping land in an agronomic trial. From the design point of view this distinction generally is not important, but it may be important with regard to dealing with possible block \times treatment interactions (see Section 9.6) and considerations of practical inference from an experiment involving blocking factors.

9.2 RANDOMIZED COMPLETE BLOCK DESIGN

9.2.1 Definition

The simplest and perhaps most widely used block design is the *randomized complete block design* (RCBD) which we define as follows: The experimental material is divided into b sets of t EUs each, where t is the number of treatments, such that the EUs within a set are as homogeneous as possible and that differences among the EUs are accounted for as much as possible by differences between the sets. The sets are called blocks. Within each block the t treatments are randomly assigned to the EUs, each treatment occurring exactly once in a block. Independent randomizations are used in the b blocks.

The physical act of randomization can be carried out for each block as described in Section 6.2.1 or by using SAS PROC PLAN as described in Table 9.1 for $t = 6$ and $b = 3$.

As alluded to above the division of the EUs into blocks is based on a priori information or what we have called the subject matter model (see Section 2.2), that is, identification of factors that may have an effect on the outcome of the experiment. It is important to identify these factors since otherwise, if only by chance, the treatments may be confounded, that is, not separable, from the “levels” of such extraneous or nuisance factors (for an example see Kempthorne, 1952).

9.2.2 Derived Linear Model

We shall now consider the analysis of data from an RCBD following the method of Kempthorne (1952, 1955) and using an approach similar to that in Chapter 6. This

Table 9.1 Randomization for Randomized Complete Block Design

a) Input statements:

```
proc plan seed=23467;
factors block=3 ordered trt=6;
title1 'RANDOMIZATION FOR RANDOMIZED COMPLETE BLOCK DESIGN';
title2 't=6, b=3';
run;
```

b) Output:

RANDOMIZATION FOR RANDOMIZED COMPLETE BLOCK DESIGN			
t=6, b=3			
The PLAN Procedure			
Factor	Select	Levels	Order
block	3	3	Ordered
trt	6	6	Random
block			
----trt----			
1	4	2	5 1 3 6
2	1	3	5 4 6 2
3	4	5	1 3 2 6

means we shall first derive a linear model and then describe the analysis of variance associated with that model. We consider first the case where EU and OU are identical. Let T_{ijk} denote the true or conceptual yield of treatment k applied to the j th EU in the i th block ($i = 1, 2, \dots, b; j = 1, 2, \dots, t; k = 1, 2, \dots, t$). We now assume, as we did in Section 6.3, treatment-unit additivity in the strict sense. Hence we write

$$T_{ijk} = U_{ij} + T_k, \quad (9.1)$$

where U_{ij} is the contribution from EU j in block i and T_k is the contribution from treatment k . Using the fact that we have formed blocks and that randomization is performed within blocks, that is, we have *restricted randomization*, we rewrite (9.1) as

$$T_{ijk} = B_i + u_{ij} + T_k, \quad (9.2)$$

where $B_i = \sum_j U_{ij}/t = \bar{U}_i$ is the average unit contribution in block i and hence referred to as the block contribution, and $u_{ij} = U_{ij} - \bar{U}_i$ with $\sum_j u_{ij} = 0$ for every i . We rewrite (9.2) further as

$$\begin{aligned} T_{ijk} &= \bar{B} + (B_i - \bar{B}) + u_{ij} + \bar{T} + (T_k - \bar{T}) \\ &= (\bar{B} + \bar{T}) + (B_i - \bar{B}) + (T_k - \bar{T}) + u_{ij} \\ &= \mu + \beta_i + \tau_k + u_{ij}, \end{aligned} \quad (9.3)$$

where the terms are defined in an obvious way. The physical interpretation of the quantities in (9.3) is as follows:

μ is the (conceptual) overall mean yield which would be obtained if each treatment were applied to every unit in every block, that is, $\mu = \bar{T} \dots$;

β_i is the difference between the (conceptual) mean yield of all treatments on all units in block i and μ , that is, $\beta_i = \bar{T}_{i..} - \bar{T} \dots$;

τ_k is the difference between the (conceptual) mean yield of treatment k applied to all units in all blocks and μ , that is, $\tau_k = \bar{T}_{..k} - \bar{T} \dots$; and

u_{ij} is the difference between the (conceptual) mean of the yields of all treatments on the j th unit of block i and the mean yield over the whole block, that is, $u_{ij} = \bar{T}_{ij.} - \bar{T}_{i..}$. It measures the extent to which unit j deviates from the other units in block i . We shall refer to this quantity (as we did to a similar quantity in Section 6.3) as the unit error (the same results following a slightly different argument were given by Wilk, 1955). It follows, of course, from the definitions that

$$\sum_i \beta_i = 0, \quad \sum_k \tau_k = 0.$$

To characterize the randomization process we introduce the design random variable $\delta_{ij}^k = 1$ if treatment k is assigned to the j th unit in block i , and $\delta_{ij}^k = 0$ otherwise. It follows then immediately that

$$P(\delta_{ij}^k = 1) = \frac{1}{t}$$

for any i, j, k because each treatment is applied to only one unit in each block. The distributional properties of the δ_{ij}^k can be derived easily following the same arguments given in Section 6.2.

Now let y_{ik} denote the observed yield of treatment k in block i . We can then write, linking the conceptual yield to the observed yield via the process of randomization,

$$y_{ik} = \sum_j \delta_{ij}^k T_{ijk}, \quad (9.4)$$

that is, the y_{ik} ($i = 1, 2, \dots, b; k = 1, 2, \dots, t$) are a realization of bt observations from the population of bt^2 conceptual observations T_{ijk} . Using (9.3) in (9.4) we obtain

$$\begin{aligned} y_{ik} &= \mu + \beta_i + \tau_k + \sum_j \delta_{ij}^k u_{ij} \\ &= \mu + \beta_i + \tau_k + \omega_{ik} \end{aligned} \quad (9.5)$$

as a derived linear model for the observations from an RCBD. The only random variable on the right-hand side of (9.5) is ω_{ik} . Its distributional properties can be established easily with the help of the distributional properties of the δ_{ij}^k . For example, we obtain

$$\begin{aligned} E_R(\omega_{ik}) &= \sum_j E_R(\delta_{ij}^k) u_{ij} \\ &= \frac{1}{t} \sum_j u_{ij} = 0 \end{aligned}$$

and

$$\begin{aligned} \text{var}_R(\omega_{ik}) &= E_R(\omega_{ik})^2 - [E_R(\omega_{ik})]^2 \\ &= E_R \left[\sum_j \delta_{ij}^k u_{ij} \right]^2 = \frac{1}{t} \sum_j u_{ij}^2 \\ &= \left(1 - \frac{1}{t}\right) \frac{\sum_j u_{ij}^2}{t-1} = \left(1 - \frac{1}{t}\right) \sigma_{iu}^2, \end{aligned} \quad (9.6)$$

where we define

$$\sigma_{iu}^2 = \frac{1}{t-1} \sum_j u_{ij}^2, \quad (9.7)$$

that is, σ_{iu}^2 measures the variability of the EUs in block i . Also, for $k \neq k'$,

$$\text{cov}_R(\omega_{ik}, \omega_{ik'}) = -\frac{1}{t} \sigma_{iu}^2, \quad (9.8)$$

that is, observations in the same block are correlated, and for $i \neq i'$

$$\text{cov}_R(\omega_{ik}, \omega_{i'k}) = 0, \quad (9.9)$$

that is, observations in different blocks are uncorrelated.

9.2.3 Estimation of Treatment Contrasts

It is obvious now that an unbiased estimator for a treatment contrast, $\sum_k c_k \tau_k$, with $\sum_k c_k = 0$, is given by the same contrast in the treatment means, that is,

$$E_R \left[\sum_k c_k \bar{y}_{..k} \right] = \sum_k c_k \tau_k$$

with

$$\text{var}_R \left[\sum_k c_k \bar{y}_{..k} \right] = \sum_k c_k^2 \text{var}_R(\bar{y}_{..k}) + \sum_{k \neq k'} c_k c_{k'} \text{cov}_R(\bar{y}_{..k}, \bar{y}_{..k'}). \quad (9.10)$$

Now

$$\begin{aligned} \text{var}_R(\bar{y}_{..k}) &= \frac{1}{b^2} \left[\sum_i \text{var}_R(\omega_{ik}) + \sum_{i \neq i'} \text{cov}_R(\omega_{ik}, \omega_{i'k}) \right] \\ &= \frac{1}{b^2} \left(1 - \frac{1}{t} \right) \sum_i \sigma_{iu}^2 \end{aligned} \quad (9.11)$$

using (9.6), (9.8), and (9.10), and

$$\begin{aligned} \text{cov}_R(\bar{y}_{..k}, \bar{y}_{..k'}) &= \frac{1}{b^2} \left[\sum_i \text{cov}_R(\omega_{ik}, \omega_{i'k'}) + \sum_{i \neq i'} \text{cov}_R(\omega_{ik}, \omega_{i'k'}) \right] \\ &= -\frac{1}{b^2} \frac{1}{t} \sum_i \sigma_{iu}^2 \end{aligned} \quad (9.12)$$

using (9.8) and (9.9). Substituting (9.11) and (9.12) into (9.10) we obtain

$$\begin{aligned} \text{var}_R \left(\sum_k c_k \bar{y}_{..k} \right) &= \frac{1}{b^2} \sum_k c_k^2 \sum_i \sigma_{iu}^2 \\ &= \frac{1}{b} \sum_k c_k^2 \frac{\sum_{ij} u_{ij}^2}{b(t-1)} \end{aligned} \quad (9.13)$$

using (9.7). To estimate (9.13) it then remains to estimate $\sum_{ij} u_{ij}^2$. This is achieved through the analysis of variance.

9.2.4 Analysis of Variance

The ANOVA table for the RCBD is obtained from the following identity, mimicking (9.5),

$$y_{ik} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.k} - \bar{y}_{..}) + (y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..}). \quad (9.14)$$

Transferring $\bar{y}_{..}$ to the left hand side of (9.14) and squaring both sides yields

$$\sum_{i,k} (y_{ik} - \bar{y}_{..})^2 = t \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + b \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2 + \sum_{i,k} (y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..})^2.$$

This is the partitioning of the total sum of squares into block, treatment, and error components as given in Table 9.2. It is easy to work out the expected values of the various sums of squares, where the expectation is taken over all possible randomizations. Since $\bar{y}_{i.} - \bar{y}_{..} = \beta_i$ is a constant (using (9.5)) we have

$$E_R[SS(B)] = SS(B) = t \sum_i \beta_i^2,$$

but $\bar{y}_{.k} - \bar{y}_{..} = \tau_k + \sum_{ij} \delta_{ij}^k u_{ij} / b$ is not a constant and hence

$$E_R[SS(T)] = b \sum_k \tau_k^2 + \frac{1}{b} \sum_{i,j} u_{ij}^2.$$

Also, $SS(\text{Total})$ is a constant and hence

$$E_R[SS(\text{Total})] = SS(\text{Total}) = t \sum_i \beta_i^2 + b \sum_k \tau_k^2 + \sum_{i,j} u_{ij}^2.$$

By subtraction we then find

$$E_R[SS(E)] = \left(1 - \frac{1}{b}\right) \sum_{i,j} u_{ij}^2.$$

From these results the $E(\text{MS})$ under additivity in the strict sense are as given in Table 9.2.

We comment briefly on these results:

- (i) We note the “asymmetry” of blocks and treatments as manifested in the different forms for $E_R[MS(B)]$ and $E_R[MS(T)]$. We shall return to this point later (see Sections 9.2.6 and 9.3).
- (ii) It follows from (9.7) that

$$\frac{1}{b(t-1)} \sum_{ij} u_{ij}^2 = \frac{1}{b} \sum_i \sigma_{iu}^2,$$

that is, the average of the variabilities of the EUs within blocks.

Table 9.2 ANOVA for RCBD

Source	d.f.	SS	E(MS)		
			MS	Strict Additivity	Broad Additivity
Blocks	$b - 1$	$t \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = SS(B)$	$MS(B)$	$t \sum_i \beta_i^2 / (b - 1)$	$\sigma_\eta^2 + \sigma_\nu^2 + t \frac{\sum_i \beta_i^2}{b - 1}$
Treatments	$t - 1$	$b \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2 = SS(T)$	$MS(T)$	$\sum_{ij} \frac{w_{ij}^2}{b(t-1)} + \frac{b \sum_k \tau_k^2}{t-1}$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2 + \frac{b \sum_k \tau_k^2}{t-1}$
Error	$(b - 1)(t - 1)$	$\sum_{i,k} (y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..})^2 = SS(E)$	$MS(E)$	$\sum_{ij} \frac{u_{ij}^2}{b(t-1)}$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2$
Total	$bt - 1$	$\sum_{i,k} (y_{ik} - \bar{y}_{..})^2$			

(iii) It follows from Table 9.2 that $\sum u_{ij}^2/b(t-1)$ is estimated by $\text{MS}(E)$ so that

$$\widehat{\text{var}}_R \left(\sum_k c_k \bar{y}_{..k} \right) = \frac{1}{b} \sum_k c_k^2 \text{MS}(E). \quad (9.15)$$

9.2.5 Randomization Test and F -Test

For testing of the null hypothesis that there are no differences among treatment effects we turn as in Sections 6.5 and 6.6 to the randomization test and its approximation by the F -test. Let us define

$$U = \sum_{ij} u_{ij}^2$$

and consider the test statistic

$$Z = \text{SS}(T)/[\text{SS}(T) + \text{SS}(E)] \quad (9.16)$$

which under $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$ is equal to

$$Z = \text{SS}(T)/U. \quad (9.17)$$

We want to compare the distribution, or more precisely the first and second moment, of Z in (9.17) under randomization theory and normal theory. Since U is a constant we have to find $E_R[\text{SS}(T)]$ and $\text{var}_R[\text{SS}(T)]$. From Table 9.2 we know that, under H_0 ,

$$E_R[\text{SS}(T)] = U/b.$$

It can be shown, after tedious and lengthy algebra (Kempthorne, 1952), that

$$\text{var}_R[\text{SS}(T)] = \frac{2}{(t-1)b^2} (U^2 - K),$$

where

$$K = \sum_i \left(\sum_j u_{ij}^2 \right)^2 = (t-1) \sum_i \sigma_{iu}^4.$$

If we assume that $\sigma_{1u}^2 = \sigma_{2u}^2 = \dots = \sigma_{bu}^2$, then $\sum_j u_{ij}^2 = U/b$ for every i and hence

$$K = U^2/b.$$

Then

$$\text{var}_R[\text{SS}(T)] = \frac{2(b-1)}{(t-1)b^3} U^2$$

and hence

$$E_R(Z) = \frac{1}{b} \quad (9.18)$$

and

$$\text{var}_R(Z) = \frac{2(b-1)}{(t-1)b^3}. \quad (9.19)$$

If the ω_{ik} 's in (9.5) were normally and independently distributed with mean zero and constant variance, then $MS(T)/MS(E)$ would follow an F -distribution with $t - 1$ and $(b - 1)(t - 1)$ d.f. or, equivalently, Z as given in (9.16) follows the beta (α, β) distribution with $\alpha = t - 1, \beta = (b - 1)(t - 1)$. It then follows (see Section 6.6) that

$$E(Z) = \frac{1}{b} \quad (9.20)$$

and

$$\text{var}(Z) = \frac{2(b - 1)(t - 1)^2}{[b(t - 1)]^2[b(t - 1) + 2]}$$

or if 2 is small compared to $b(t - 1)$,

$$\text{var}(Z) \approx \frac{2(b - 1)}{b^3(t - 1)}. \quad (9.21)$$

It then follows from comparing (9.18) with (9.20) and (9.19) with (9.21) that the means of Z are the same under randomization theory and normal theory and that the variances are approximately equal. We conclude from this that the F -test

$$F = \frac{MS(T)}{MS(E)} \quad (9.22)$$

is a reasonable approximation to the randomization test to test $H_0: \tau_1 = \tau_2 = \cdots = \tau_t = 0$, a result first obtained by Welch (1937) and Pitman (1937) following Fisher (1935). Just as in Chapter 6 this result can be further substantiated by computational methods, either by enumerating all possible randomizations or, if that proves to be prohibitive, through simulation (Monte Carlo) studies.

Individual contrasts among treatment effects are tested by using a t -test (or equivalent tests for multiple comparisons; see Chapter 7) in connection with result (9.15). There is no theoretical justification for this approximation but empirical results, such as simulation studies, seem to indicate that such a procedure yields satisfactory results (see Kempthorne and Doerfler, 1969).

9.2.6 Additivity in the Broad Sense

Up to this point we have considered the case where the assumption of additivity in the strict sense holds. It is, of course, desirable and indeed necessary to broaden our assumptions and extend model (9.5) so as to include in addition to unit error, other errors such as treatment error and observational (sampling) error. This can be done by using the same arguments as given in Section 6.3. We shall not give details here except to state the model under the assumption of additivity in the broad sense as

$$y_{ik} = \mu + \beta_i + \tau_k + \omega_{ik} + \nu_{ik} + \eta_{ik}, \quad (9.23)$$

where ν_{ik} is the treatment error and η_{ik} is the observational error with means zero and variances σ_ν^2 and σ_η^2 , respectively. Using this model in the ANOVA table (Table 9.2)

the $E(\text{MS})$ can be obtained easily and are as given in the right-hand column under $E(\text{MS})$. If we define

$$\sum_{ij} u_{ij}^2 / b(t-1) = \sigma_u^2$$

we recognize the similarity of the $E(\text{MS})$ under models (9.5) and (9.23) in that for $E[\text{MS}(T)]$ and $E[\text{MS}(E)]$, σ_u^2 has been replaced by $\sigma_u^2 + \sigma_v^2 + \sigma_\eta^2 = \sigma_e^2$, say, whereas for $E[\text{MS}(B)]$ only $\sigma_v^2 + \sigma_\eta^2$ has been added. Thus, under model (9.23) the “asymmetry” between blocks and treatments discussed earlier is still preserved. This implies three things:

- (i) Under H_0 : $\tau_1 = \tau_2 = \cdots = \tau_t$, $\text{MS}(T)$ and $\text{MS}(E)$ have the same expected value, that is, the design possesses the property of unbiasedness;
- (ii) assuming equality of the unit error variances the statistic (9.22) can still be used to test $H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$;
- (iii) there does not exist a valid test for testing equality of block effects (we shall return to this point in Section 9.3).

We shall elaborate briefly on the result (iii) above since we consider this to be an important and often not understood finding. It formalizes what should be intuitively obvious, namely that a distinction needs to be made between interventional and observational studies in general, and the RCBD and the two-factor observational study, specifically. With regard to the latter, in both situations the observations are expressed in terms of a two-way classificatory linear model (see Section 4.3.2). For the observational study the two factors in this model are equivalent (symmetric), whereas for the experimental study they are asymmetric: the treatments (levels of factor A) are randomly assigned to the EUs, but the blocks (levels of factor B) are not randomly assigned. That this should lead to different properties – related to statistical inference – for the treatment and block effects of model (9.23) becomes explicit only through careful consideration of the various error components as exhibited in (9.23) and subsequent application of randomization theory. This is in sharp contrast to the usual – and incorrect, we might add – discussion of this important and far reaching topic.

To conclude and relate this discussion to that in Section 6.3 we note that $\varepsilon_{ik} = \omega_{ik} + \nu_{ik}$ in (9.23) is referred to as the experimental error and that for all practical purposes, that is, for purposes of inferences about treatment contrasts, the ε_{ik} can be regarded as i.i.d. random variables with mean zero and variance σ_e^2 . We may write further $\varepsilon_{ik} + \eta_{ik} = e_{ik}$, and hence model (9.23) as

$$y_{ik} = \mu + \beta_i + \tau_k + e_{ik},$$

where the e_{ik} can be considered also as i.i.d. random variables with mean zero and variance

$$\sigma_e^2 = \sigma_\varepsilon^2 + \sigma_\eta^2.$$

We then have for a contrast of treatment means, $\sum c_k \bar{y}_{.k}$,

$$\text{var} \left(\sum c_k \bar{y}_{.k} \right) = \frac{1}{b} \sum_k c_k^2 \sigma_e^2 \quad (9.24)$$

and tests can be made in the familiar way.

9.2.7 Subsampling in an RCBD

Just as in a CRD (see Section 6.9) we can encounter in an RCBD (as, in fact, in any error-control design) the situation that EUs and OUs are not identical. For an illustration consider Example 2: Experimental Situation IV in Table 2.6. We refer to this as an *RCBD with subsampling*. The important point here is that now model (9.23) can be written as

$$y_{ikl} = \mu + \beta_i + \tau_k + \omega_{ik} + \nu_{ik} + \eta_{ikl}$$

or

$$y_{ikl} = \mu + \beta_i + \tau_k + \epsilon_{ik} + \eta_{ikl}$$

with $l = 1, 2, \dots, n$ and n indicating the number of OUs for each EU. As a consequence we are able to estimate the experimental error variance component $\sigma_\epsilon^2 = \sigma_u^2 + \sigma_v^2$ and observational error variance component σ_v^2 separately. This follows in an obvious way from the expected mean squares in Table 9.3, namely,

$$\hat{\sigma}_v^2 = \text{MS}(OE)$$

$$\hat{\sigma}_\epsilon^2 = [\text{MS}(EE) - \text{MS}(OE)]/n.$$

And, most importantly, the hypothesis $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$ can be tested by approximating the randomization test by the F -test

$$F = \frac{\text{MS}(T)}{\text{MS}(EE)}$$

with $t-1$ and $(b-1)(t-1)$ d.f. Also, since for a contrast of treatment means, $\sum c_k \bar{y}_{.k}$, with $\sum c_k = 0$ we find

$$\text{var} \left(\sum_k c_k \bar{y}_{.k} \right) = \frac{1}{bn} \sum_k c_k^2 (\sigma_\eta^2 + n\sigma_\epsilon^2)$$

and $\sigma_\eta^2 + n\sigma_\epsilon^2$ is estimated by $\text{MS}(EE)$ it should be clear that $\text{MS}(EE)$ plays now the important role in any inference concerning the treatment effects.

9.3 RELATIVE EFFICIENCY OF THE RANDOMIZED COMPLETE BLOCK DESIGN

9.3.1 Question of Effectiveness of Blocking

In many practical situations it is quite obvious that there are substantial differences between the blocks, and hence between the block effects, that is, the B_i 's in terms of model (9.2) or the β_i 's in terms of model (9.3). In such cases there is no doubt that the naturally arising or created blocks should be utilized for purposes of reducing experimental error, leading to a "small" σ_u^2 . There are, however, situations where matters

Table 9.3 ANOVA for RCBD with Subsampling

Source	d.f.	SS	E(MS)
Blocks	$b - 1$	$tn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = SS(B)$	—
Treatments	$t - 1$	$bn \sum_k (\bar{y}_{.k.} - \bar{y}_{...})^2 = SS(T)$	$\sigma_\eta^2 + n\sigma_\epsilon^2 + \frac{bn \sum_k \tau_k^2}{t - 1}$
Exp. Error	$(b - 1)(t - 1)$	$n \sum_{i,k} (\bar{y}_{ik.} - \bar{y}_{i..} - \bar{y}_{.k.} + \bar{y}_{...})^2 = SS(EE)$	$\sigma_\eta^2 + n\sigma_\epsilon^2$
Obs. Error	$bt(n - 1)$	$\sum_{i,k,l} (\bar{y}_{ikl} - \bar{y}_{ik.})^2 = SS(OE)$	σ_η^2
Total	$btn - 1$	$\sum_{i,k,l} (y_{ikl} - \bar{y}_{...})^2$	

are not as clear or only incomplete information about the blocks is available at the outset of the experiment. The investigator, relying on a subject matter model (see Chapter 2), may have used an ineffective blocking factor, that is, a blocking factor that leads to only a small reduction in error compared to a CRD. This reduction in error may, however, be offset by a loss of d.f. for $SS(E)$ for the same number of observations, that is, $t(r-1)$ for the CRD versus $(t-1)(r-1)$ for the RCBD with $b = r$ blocks. This then may result in a loss of power or sensitivity with respect to treatment comparisons.

Even though, once an experiment has been conducted in a RCBD, one cannot ignore the blocking in the analysis one may ask the question: How much have we gained by using a RCBD rather than a CRD with the same number of experimental units? To answer this question may be useful if one were to conduct a similar experiment using the same or similar EUs in the future.

To study this question Yates (1935) introduced the notion of *relative efficiency* (RE) in the context of estimation of treatment comparisons. For two designs, D_1 , and D_2 say, the RE of D_1 to D_2 is defined as

$$\begin{aligned} \text{RE}(D_1 \text{ to } D_2) &= \frac{\text{Efficiency } D_1}{\text{Efficiency } D_2} \\ &= \frac{\text{var}_{D_2}}{\text{var}_{D_1}}, \end{aligned} \quad (9.25)$$

where var_{D_l} refers to $\text{var}(\sum c_k \hat{\tau}_k)$ for design D_l ($l = 1, 2$). In our case D_1 is a RCBD with t treatments and r blocks and D_2 is a CRD with r replications for each of the t treatments. The RE as defined in (9.25) depends on the true variances for the two designs which, of course, are unknown. The best we can do then is to obtain the estimated RE, which we shall denote by ERE. Moreover, we have available only the data (observations) from the RCBD.

9.3.2 Use of Uniformity Trials

Following Yates (1935) we consider a uniformity trial, that is, a trial with dummy treatments, with b blocks and t EUs in each block. Denote the observations from such a trial by y_{ij} ($i = 1, 2, \dots, b; j = 1, 2, \dots, t$). The ANOVA table for data with such a structure is as given in Table 9.4.

If the blocks were not used the estimated error variance would be

$$\frac{SS(B) + SS(R)}{bt - 1} = \frac{(b-1)MS(B) + b(t-1)MS(R)}{bt - 1}.$$

The estimated error variance with blocks is, of course, $MS(R)$, so that

$$\text{ERE(RCBD to CRD)} = \frac{(b-1)MS(B) + b(t-1)MS(R)}{(bt-1)MS(R)}. \quad (9.26)$$

Since we have carried out an experiment with real treatments and not dummy treatments we do not know $MS(R)$. Instead we only know $MS(E)$ from Table 9.2. Hence substituting $MS(E)$ for $MS(R)$ in (9.26) yields

$$\text{ERE(RCBD to CRD)} = \frac{(b-1)MS(B) + b(t-1)MS(E)}{(bt-1)MS(E)}, \quad (9.27)$$

Table 9.4 ANOVA for Uniformity Trial

Source	d.f.	SS	MS
Blocks	$b - 1$	$SS(B)$	$MS(B)$
Within Blocks (Error)	$b(t - 1)$	$SS(R)$	$MS(R)$
Total	$bt - 1$	$SS(B) + SS(R)$	

where $MS(B)$ is also taken from Table 9.2. It is useful to mention that (9.27) can also be obtained by randomization arguments only, that is, by comparing restricted (RCBD) versus unrestricted (CRD) randomization (Kempthorne, 1955).

To conclude this discussion we should mention that sometimes it may be quite feasible and appropriate to conduct a uniformity trial before actually using an RCBD. In that case we then know $MS(R)$ from Table 9.4 and hence can use (9.26) to obtain the ERE.

9.3.3 Interpretation and Use of Relative Efficiency

In general $MS(B)$ will be larger than $MS(E)$ and hence ERE will be larger than one or, as it is usually presented, larger than 100%. Since the ERE is obtained when both the RCBD and CRD have the same number of replications, namely, b , expression (9.27) can then be rewritten as

$$ERE = \frac{\widehat{\text{var}}_{CRD}/b}{\widehat{\text{var}}_{RCBD}/b}$$

or

$$\frac{\widehat{\text{var}}_{CRD}}{b \cdot ERE} = \frac{\widehat{\text{var}}_{RCBD}}{b}.$$

The practical interpretation of ERE thus is that we require

$$r = b \times ERE \quad (9.28)$$

replications per treatment for a CRD to be as effective as the RCBD with b replications, that is, b blocks, using the same experimental material. We emphasize again that the ERE speaks only to the question of estimation, that is, precision of estimates, and not to the question of power, that is, sensitivity of the experiment. For this reason it may be advisable to consider only a RCBD with an ERE larger than, say, 125% to be “better” than the comparable CRD. Another interpretation of the ERE is given by Yates (1935), namely that $(1 - 1/ERE)$ 100% is the percent variation among the EUs removed by blocking.

We commented earlier (Section 9.2) that there does not exist a legitimate test for $H_0: \beta_1 = \beta_2 = \cdots = \beta_b$, at least not in the context of the ANOVA table, that is, $H = MS(B)/MS(E)$ is not an appropriate test statistic. There exists, however, a

monotonic relationship between H and ERE which at least gives some meaning to H (Lentner, Arnold, and Hinkelmann, 1989). It follows from (9.27) that

$$\text{ERE} = \alpha + (1 - \alpha) H,$$

where $\alpha = b(t - 1)/(bt - 1)$. Hence $\text{ERE} > 1$ if and only if $H > 1$. This gives a certain usefulness to H , but referring to our earlier discussion, it tells only part of the story.

Finally, some knowledge of ERE or $\text{ERE}^* = 1 - 1/\text{ERE}$ may be useful in determining the number of blocks b to be used if one has determined r for a comparable CRD by using the tables of Bowman and Kastenbaum (1975) as discussed in Section 6.7. We have from (9.28) that

$$b = \frac{r}{\text{ERE}} = r(1 - \text{ERE}^*).$$

For example, if we have some idea that blocking will reduce variability by 25% then 25% fewer replications than those required for the CRD are necessary. Bowman and Kastenbaum (1975) provide a limited set of tables for the number of blocks for the RCBD, but the above procedure may be quite satisfactory from a practical point of view.

9.4 SUPPLEMENTARY INFORMATION AND ANALYSIS OF COVARIANCE

9.4.1 The Model

One method of generating blocks is to make use of supplementary information in the form of a covariate. The procedure is to rank the EUs with respect to the covariate (which of course must be available before the experiment) in increasing order of magnitude, often referred to as outcome groups, and then use the first t EUs as one block, the next t EUs as another block, and so on. Cox (1957) has shown that this method is preferable to a CRD with covariate unless the correlation between y and x is at least .6. Using the same covariate for purposes of analysis in addition to its use as a blocking device will generally not provide much additional information. There may, however, be situations where in addition to blocking the use of some covariate will lead to further reduction of the error variance, that is, we may consider the model

$$y_{ik} = \mu + \beta_i + \tau_k + \gamma(x_{ik} - \bar{x}_{..}) + e_{ik}^*, \quad (9.29)$$

where the e_{ik}^* can be considered as i.i.d. random variables with mean 0 and variance σ_e^2 . This is, of course, an obvious extension of the analysis of covariance procedure discussed in Chapter 8. We shall now give a brief description of the technique for the RCBD without repeating the basic philosophy and assumptions set forth in Chapter 8. This discussion also serves as an example of extending this technique to other error-reduction designs as well (for a general discussion of the arithmetic of analysis of covariance see also Section 4.13).

9.4.2 Least Squares Analysis

Model (9.29) is a special case of model (8.55)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}^*, \quad (9.30)$$

where $\mathbf{X}\boldsymbol{\mu}$ represents the classificatory part

$$(\mathcal{J}|\mathbf{X}_\beta|\mathbf{X}_\tau) \begin{bmatrix} \mu \\ \beta \\ \tau \end{bmatrix}$$

for the RCBD with

$$\mathbf{X}_\beta = \begin{bmatrix} \mathcal{J}_t & & & \\ & \mathcal{J}_t & & \\ & & \ddots & \\ & & & \mathcal{J}_t \end{bmatrix}_{bt \times b}$$

and

$$\mathbf{X}_\tau = \begin{bmatrix} \mathbf{I}_t \\ \mathbf{I}_t \\ \vdots \\ \mathbf{I}_t \end{bmatrix}_{bt \times t}$$

Then, using model (8.58) we have in the normal equations (8.60)

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} bt & t\mathcal{J}'_b & b\mathcal{J}'_t \\ t\mathcal{J}_b & t\mathbf{I}_b & \mathcal{J}_b\mathcal{J}'_t \\ b\mathcal{J}_t & \mathcal{J}_t\mathcal{J}'_b & b\mathbf{I}_t \end{bmatrix}$$

with $\text{rank}(\mathbf{X}'\mathbf{X}) = b + t - 1$. A g -inverse of $\mathbf{X}'\mathbf{X}$ is obtained by imposing the conditions $\Sigma\hat{\beta}_i = 0, \Sigma\hat{\tau}_k = 0$. Corresponding to (8.61) we then have

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{(0)} &= (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y} \\ &= \begin{bmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{b.} - \bar{y}_{..} \\ \bar{y}_{.1} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{.t} - \bar{y}_{..} \end{bmatrix}. \end{aligned}$$

Furthermore, because of the definition of $\mathbf{R}_\mathbf{X}$ as the matrix for the error sums of squares, we obtain $\hat{\gamma}$ from (8.62) as

$$\hat{\gamma} = \frac{E_{xy}}{E_{xx}}, \quad (9.31)$$

where E_{xy} and E_{xx} are the error sum of products for x and y and the error sum of squares for x for the RCBD, respectively. Finally, we have corresponding to (8.63), for example

$$\hat{\tau}_k = \bar{y}_{..k} - \bar{y}_{..} - \hat{\gamma}(\bar{x}_{..k} - \bar{x}_{..}) \quad (9.32)$$

or

$$\hat{\mu} + \hat{\tau}_k = \bar{y}_{..k} - \hat{\gamma}(\bar{x}_{..k} - \bar{x}_{..}).$$

All the other results follow similarly.

9.4.3 The ANOVA Table

We shall comment now briefly on the ANOVA table. It follows from (8.66) that the error sum of squares can be written as

$$SS(\mathbf{I}|\mathbf{X}, \mathbf{Z}) = E_{yy} - \frac{E_{xy}^2}{E_{xx}}, \quad (9.33)$$

where, again, E_{yy} , E_{xx} , E_{xy} refer to error sums of squares and error sum of products, respectively, for the RCBD. Hence

$$\hat{\sigma}_{e^*}^2 = \left[E_{yy} - \frac{E_{xy}^2}{E_{xx}} \right] / [(b-1)(t-1) - 1]. \quad (9.34)$$

To test the hypothesis $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$, we obtain the $SS(\text{Treatments})$ as

$$SS(\mathbf{X}_\tau | \mathbf{J}, \mathbf{X}_\beta, \mathbf{Z}) = SS(\mathbf{I}|\mathbf{X}^*, \mathbf{Z}) - SS(\mathbf{I}|\mathbf{X}, \mathbf{Z}),$$

where

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{J}_t & & & \\ & \mathbf{J}_t & & \\ & & \ddots & \\ \mathbf{J}_{bt} & & & \mathbf{J}_t \end{bmatrix}_{bt \times (1+b)}$$

or, if we write $\mu + \beta_i = \mu_i$,

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{J}_t & & & \\ & \mathbf{J}_t & & \\ & & \ddots & \\ & & & \mathbf{J}_t \end{bmatrix}_{bt \times b}.$$

It then follows from the results for the CRD that

$$SS(\mathbf{I}|\mathbf{X}^*, \mathbf{Z}) = (T_{yy} + E_{yy}) - \frac{(T_{xy} + E_{xy})^2}{T_{xx} + E_{xx}} \quad (9.35)$$

and hence, using (9.35) and (9.33)

$$SS(\mathbf{X}_\tau | \mathbf{J}, \mathbf{X}_\beta, \mathbf{Z}) = T_{yy} - \frac{(T_{xy} + E_{xy})^2}{T_{xx} + E_{xx}} + \frac{E_{xy}^2}{E_{xx}}. \quad (9.36)$$

The test statistic for H_0 is then obtained by substituting (9.36) and (9.34) in (8.68) with $d = t - 1$.

The reader should recognize that the right-hand sides of (9.36) and (8.28) are of the exact same form since for the CRD we have $S_{xy} = T_{xy} + E_{xy}$ and $S_{xx} = T_{xx} + E_{xx}$ see Table 8.1. Moreover, (9.31), (9.32), and (9.36) carry over to all other error-control designs in the following chapters keeping in mind only that E_{yy} , E_{xx} , and E_{xy} are the appropriate error sums of squares and products, respectively, for the error-control design under consideration.

9.5 MISSING OBSERVATIONS

Even in well-planned experiments it may happen that, for reasons that cannot be ascribed to the effect of the treatments, one or several observations may not be available. This destroys the simplicity of the analysis of such data, but unless the missing observations occur in a particular pattern the experiment is not a complete failure. With existing statistical software such data can be handled easily on any computer. Using a general linear models program the least squares analysis can be performed and all necessary information will be provided. In essence then the design becomes an incomplete block design and methods for dealing with such designs are described explicitly in Chapter II.1.

Historically, this topic has received a great deal of attention, mainly for purely computational reasons. Yates (1933) developed a procedure for estimating missing observations, substituting the estimates for the missing observations and then analyzing the thus completed data set in the usual fashion. This leads to an approximate analysis which, however, is quite satisfactory for most purposes. As we mentioned above, there is today no particular reason to describe and use this method for the RCBD. Yet we shall describe a particular method of estimating missing values here for the following reasons: (i) It may not always be possible to perform the least squares analysis from first principles for complex and highly structured data sets because the large number of parameters leads to normal equations which cannot be solved on existing computers (see for example, Perry, 1986); (ii) the method to be described for estimating missing observations is generally applicable but easily described and illustrated for the RCBD; and (iii) the method is applicable to situations other than experimental designs (see Hinkelmann, 1968).

9.5.1 Estimating a Missing Observation

The method we shall describe now was originally proposed by M. S. Bartlett and expanded by Coons (1957) and is based on analysis of covariance techniques (see Chapter 8). Consider then a RCBD with t treatments in b blocks and suppose that the observation for treatment k^* in block i^* is missing. We then write the model for the observations from this design as

$$y_{ik} = \mu + \beta_i + \tau_k + \gamma x_{ik} + e_{ik}, \quad (9.37)$$

where

$$\begin{aligned} y_{ik} &= \begin{cases} 0 & \text{for } i = i^*, \quad k = k^* \\ y_{ik} & \text{otherwise} \end{cases} \\ x_{ik} &= \begin{cases} -1 & \text{for } i = i^*, \quad k = k^* \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

It follows then from (9.37) immediately that $\hat{\gamma}$ is an estimate of $y_{i^*k^*}$ and from (9.31) we have

$$\hat{\gamma} = \frac{E_{xy}}{E_{xx}}. \quad (9.38)$$

Now, using the special nature of y_{ik} and x_{ik} above,

$$\begin{aligned} E_{xy} &= \sum_{ik} (x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x}_{..})(y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..}) \\ &= \sum_{ik} x_{ik}y_{ik} - t \sum_i \bar{x}_{i.}\bar{y}_{i.} - b \sum_k \bar{x}_{.k}\bar{y}_{.k} + bt\bar{x}_{..}\bar{y}_{..} \end{aligned}$$

and

$$E_{xx} = \sum_{ik} x_{ik}^2 - t \sum_i \bar{x}_{i.}^2 - b \sum_k \bar{x}_{.k}^2 + bt\bar{x}_{..}^2.$$

Substituting the values for x_{ik} and y_{ik} as defined above we obtain

$$\begin{aligned} E_{xy} &= 0 + \bar{y}_{i^*} + \bar{y}_{.k^*} - \bar{y}_{..} \\ &= \frac{1}{t}B_{i^*} + \frac{1}{b}T_{k^*} - \frac{1}{bt}G, \end{aligned} \quad (9.39)$$

where

B_{i^*} = total of all observations in block i^*

T_{k^*} = total of all observations for treatment k^*

G = grand total

and

$$\begin{aligned} E_{xx} &= 1 - \frac{1}{b} - \frac{1}{t} + \frac{1}{bt} \\ &= \frac{(b-1)(t-1)}{bt}. \end{aligned} \quad (9.40)$$

Substituting (9.39) and (9.40) into (9.38) yields

$$\hat{\gamma} = \frac{bB_{i^*} + tT_{k^*} - G}{(b-1)(t-1)}.$$

We can then substitute $\hat{\gamma}$ for the missing observation and proceed with the analysis of the RCBD as outlined in Section 9.2.

9.5.2 Using the Estimated Missing Observation

As a consequence of the procedure advocated in Section 9.5.1 we have the following:

- (i) The comparison between treatment k^* and any other treatment k is given by

$$\hat{\tau}_{k^*} - \hat{\tau}_k = \frac{1}{b}(T_{k^*} + \hat{\gamma}) - \bar{y}_{.k}.$$

- (ii) The same result is obtained by using the analysis of covariance model (9.37) and the procedure described in Chapter 8, namely,

$$\hat{\tau}_{k^*} - \hat{\tau}_k = \bar{y}_{.k^*} - \bar{y}_{.k} - \hat{\gamma}(\bar{x}_{.k^*} - \bar{x}_{.k})$$

with

$$\bar{x}_{.k^*} = -\frac{1}{b}, \quad \bar{x}_{.k} = 0 \quad (k \neq k^*).$$

- (iii) It follows easily from the results of Chapter 8 that

$$\begin{aligned} \text{var}(\hat{\tau}_{k^*} - \hat{\tau}_k) &= \left(\frac{2}{b} + \frac{1}{b^2} \frac{bt}{(b-1)(t-1)} \right) \sigma_e^2 \\ &= \left(\frac{2}{b} + \frac{t}{b(b-1)(t-1)} \right) \sigma_e^2. \end{aligned}$$

- (iv) For $k, k' \neq k^*$

$$\begin{aligned} \hat{\tau}_k - \hat{\tau}_{k'} &= \bar{y}_{.k} - \bar{y}_{.k'} \\ \text{var}(\hat{\tau}_k - \hat{\tau}_{k'}) &= \frac{2}{b} \sigma_e^2. \end{aligned}$$

- (v)

$$\text{SS}(E) = E_{yy} - \frac{E_{xy}^2}{E_{xx}},$$

where E_{yy} is obtained with $y_{i^*k^*} = 0$ and E_{xy} and E_{xx} are as given in (9.39) and (9.40), and

$$\text{MS}(E) = \text{SS}(E)/[(b-1)(t-1)-1] = \hat{\sigma}_{e^*}^2, \quad (9.41)$$

that is, the d.f. are reduced by one for the one missing observation. We note here that $\text{MS}(E)$ in (9.41) is the same as would be obtained from the least squares analysis for the incomplete data set (see also Chapter II.1).

- (vi) The $\text{SS}(T)$ with $\hat{\gamma}$ substituted for $y_{i^*k^*}$ is positively biased (see Exercise 9.10) and hence the usual F -test for testing $H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$ will only be approximate. Only in borderline cases of significance, however, will one need to obtain the exact $\text{SS}(T)$ from the least squares analysis (see Chapter II.1) or correct for the bias (as found in Exercise 9.10).

9.5.3 Several Missing Observations

If two observations or more are missing we can extend model (9.37) and include one covariate for each missing observation. The methods of Section 8.7 can then be employed using the same ideas as outlined above. Explicit missing value formulas are given by Glenn and Kramer (1958).

We shall conclude this section with a brief discussion of the general case, that is, the case of m missing values and error-control designs more general than the RCBD (for example, the Latin square design of Chapter 10). To facilitate the discussion we use notation of Chapter 4 (in particular Section 4.13).

Let

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e} \quad (9.42)$$

represent the model for a given error-control design where β represents all the parameters associated with that design. For the RCBD, for example, β represents the constant μ , the block effects $\beta_1, \beta_2, \dots, \beta_b$, and the treatment effects $\tau_1, \tau_2, \dots, \tau_t$. Let us now write (9.42) as

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \beta + \mathbf{e} \quad (9.43)$$

and let us suppose that \mathbf{y}_1 is being observed and \mathbf{y}_2 , representing m observations, is missing. Corresponding to (9.37) we then introduce the model

$$\mathbf{y}^* = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \beta + \mathbf{Z}\gamma + \mathbf{e}, \quad (9.44)$$

that is, we replace in (9.43) the vector of the missing observations, \mathbf{y}_2 , by $\mathbf{0}$ and introduce covariates in the form of the matrix

$$\mathbf{Z} = \begin{pmatrix} \emptyset \\ -\mathbf{I} \end{pmatrix}$$

with $\mathbf{I} = \mathbf{I}_m$. The NE for model (9.44) are then

$$\begin{aligned} \mathbf{X}'\mathbf{X}\hat{\beta} - \mathbf{X}'_2\hat{\gamma} &= \mathbf{X}'\mathbf{y}^* = \mathbf{X}'_1\mathbf{y}_1 \\ -\mathbf{X}_2\hat{\beta} + \hat{\gamma} &= \mathbf{Z}'\mathbf{y}^* = \mathbf{0}. \end{aligned} \quad (9.45)$$

By the usual covariance argument (see Section 4.13) we obtain from (9.45)

$$\mathbf{Z}'[\mathbf{I} - \mathbf{P}_\mathbf{X}]\mathbf{Z}\hat{\gamma} = \mathbf{Z}'[\mathbf{I} - \mathbf{P}_\mathbf{X}]\mathbf{y}^*. \quad (9.46)$$

We recognize, of course, that the elements of the coefficient matrix on the LHS of (9.46) are obtained as the error sums of squares and error sums of products for the given error-control design with the m columns of \mathbf{Z} used as “observation” vectors. Similarly, the RHS of (9.46) is obtained as the error sums of products of the columns of \mathbf{Z} and the vector \mathbf{y}^* as defined in (9.44). Solutions to the equations (9.46) for the RCBD are the missing value formulas given by Glenn and Kramer (1958) mentioned earlier.

Now let

$$\mathbf{I} - \mathbf{P}_X = \begin{pmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{pmatrix},$$

where $\mathbf{U}, \mathbf{V}, \mathbf{W}$ are of order $(n-m) \times (n-m)$, $(n-m) \times m$, and $m \times m$, respectively, where n is the size of \mathbf{y}^* . Since $\mathbf{I} - \mathbf{P}_X$ is idempotent we have

$$\begin{pmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{pmatrix} \begin{pmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{pmatrix} = \begin{pmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W} \end{pmatrix}$$

and hence $\mathbf{V}'\mathbf{V} + \mathbf{W}^2 = \mathbf{W}$. We also have

$$\begin{aligned} \mathbf{Z}'[\mathbf{I} - \mathbf{P}_X]\mathbf{y}^* &= -\mathbf{V}'\mathbf{y}_1 \\ \mathbf{Z}'[\mathbf{I} - \mathbf{P}_X]\mathbf{Z} &= \mathbf{V}'\mathbf{V} + \mathbf{W}^2 = \mathbf{W}. \end{aligned}$$

The NE (9.46) for γ then are

$$\mathbf{W}\hat{\gamma} = -\mathbf{V}'\mathbf{y}_1$$

and hence

$$\hat{\gamma} = -\mathbf{W}^{-1}\mathbf{V}'\mathbf{y}_1. \quad (9.47)$$

It follows then from (9.47) that

$$\begin{aligned} \text{var}(\hat{\gamma}) &= \mathbf{W}^{-1}\mathbf{V}'\mathbf{V}\mathbf{W}^{-1}\sigma_e^2 \\ &= \mathbf{W}^{-1}(\mathbf{W} - \mathbf{W}^2)\mathbf{W}^{-1}\sigma_e^2 \\ &= (\mathbf{W}^{-1} - \mathbf{I})\sigma_e^2. \end{aligned}$$

Returning now to the NE for β we have from (9.45) and (9.47)

$$\begin{aligned} \mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{X}'_1\mathbf{y}_1 + \mathbf{X}'_2\hat{\gamma} \\ &= \mathbf{X}'_1\mathbf{y}_1 - \mathbf{X}'_2\mathbf{W}^{-1}\mathbf{V}'\mathbf{y}_1. \end{aligned} \quad (9.48)$$

Since the model value of the LHS of (9.48) equals the model value of the RHS we have

$$\mathbf{X}'\mathbf{X} = \mathbf{X}'_1\mathbf{X}_1 - \mathbf{X}'_2\mathbf{W}^{-1}\mathbf{V}'\mathbf{X}_1$$

and

$$\mathbf{X}'_2\mathbf{W}^{-1}\mathbf{V}'\mathbf{X}_1 = -[\mathbf{X}'\mathbf{X} - \mathbf{X}'_1\mathbf{X}_1] = \mathbf{X}'_2\mathbf{X}_2. \quad (9.49)$$

It follows then from (9.48) and using (9.49) that

$$\begin{aligned} \text{var}(\mathbf{X}'\mathbf{X}\hat{\beta}) &= (\mathbf{X}'_1 - \mathbf{X}'_2\mathbf{W}^{-1}\mathbf{V}')(\mathbf{X}_1 - \mathbf{V}\mathbf{W}^{-1}\mathbf{X}_2)\sigma_e^2 \\ &= [\mathbf{X}'\mathbf{X} + \mathbf{X}'_2\mathbf{W}^{-1}\mathbf{X}_2]\sigma_e^2. \end{aligned} \quad (9.50)$$

In the context of our approach to the analysis of data from designed experiments, (9.50) should be used to obtain the variances of estimated treatment contrasts. It shows that if a treatment contrast does not involve a treatment with missing observations then the variance is the same as would have been obtained from the complete design. For other treatment contrasts, (9.50) shows how the variance will have to be adjusted, that is, increased.

9.6 NONADDITIVITY IN THE RCBD

9.6.1 The Problem of Nonadditivity

In our discussion of the RCBD so far we have made the assumption of treatment-unit additivity in the strict sense [see (9.1) and (9.3)] or, more realistically, additivity in the broad sense [see (9.23)]. In most cases such an assumption is not unrealistic, but we may, of course, conceive of situations where it does not hold. Such a situation we shall refer to as *nonadditivity*. An explicit formulation of this situation can be given by amending model (9.1) to

$$T_{ijk} = U_{ij} + T_k + R_{ijk}, \quad (9.51)$$

where R_{ijk} refers to the nonadditivity which we also call *unit-treatment interaction*. Such interaction may arise in two ways:

- (i) The effect of the treatment depends on the EU to which it is applied in the sense that if we could apply two treatments, say k and k' , to the same EU, then we would observe

$$T_{ijk} - T_{ijk'} \neq T_{ij'k} - T_{ij'k'}$$

for two EUs j and j' in block i .

- (ii) The effect of the treatment depends on the block in which it is applied in the sense that if the EUs in a block were identical

$$T_{ijk} - T_{ij'k'} \neq T_{ij''k} - T_{ij'''k'}$$

for any two EUs in blocks i and i' .

We may refer to the first type of interaction as the *strict unit-treatment interaction*, and to the second as *block-treatment interaction*.

It should be clear from the description of the nature of strict unit-treatment interaction that there is no way to investigate whether it exists (see Kempthorne, 1952, and Wilk, 1955) because in the RCBD we can apply only one treatment to each EU. Even the block-treatment interaction, and it surely is the more important of the two, cannot always be addressed satisfactorily. We shall distinguish here between two situations: (i) there is only one blocking factor, either a non-specific factor (\mathcal{U}) or an intrinsic factor (\mathcal{Z}); (ii) there are several blocking factors involving factors from \mathcal{Z} and \mathcal{U} . For scenario (i) we shall describe two ad hoc procedures (Sections 9.6.3 9.6.5), and for scenario (ii) we shall outline appropriate analysis of variance procedures (Section 9.6.7). An alternate procedure, addressing design questions, will be discussed in Section 9.7.

9.6.2 General Model for Nonadditivity

In light of our discussion above, we can rewrite (9.51) as

$$T_{ijk} = U_{ij} + T_k + Q_{ik} + S_{ijk}, \quad (9.52)$$

where Q_{ik} represents the block-treatment interaction and S_{ijk} the strict unit-treatment interaction. Following (9.2) and (9.3), model (9.52) can be written as

$$\begin{aligned} T_{ijk} &= B_i + u_{ij} + T_k + Q_{ik} + S_{ijk} \\ &= \mu + \beta_i + \tau_k + (\beta\tau)_{ik} + u_{ij} + s_{ijk}, \end{aligned}$$

the terms of which are defined as functions of the T_{ijk} as follows (see Wilk, 1955)

$$\begin{aligned} \mu &= \bar{T}_{...} \\ \beta_i &= \bar{T}_{i..} - \bar{T}_{...} \\ \tau_k &= \bar{T}_{..k} - \bar{T}_{...} \\ (\beta\tau)_{ik} &= (\bar{T}_{i.k} - \bar{T}_{i..}) - (\bar{T}_{..k} - \bar{T}_{...}) \\ u_{ij} &= \bar{T}_{ij.} - \bar{T}_{i..} \\ s_{ijk} &= (T_{ijk} - \bar{T}_{ij.}) - (\bar{T}_{i.k} - \bar{T}_{i..}). \end{aligned}$$

The actual observations, y_{ik} , can then be expressed again as

$$\begin{aligned} y_{ik} &= \sum_j \delta_{ij}^k T_{ijk} \\ &= \mu + \beta_i + \tau_k + (\beta\tau)_{ik} + \sum_j \delta_{ij}^k u_{ij} + \sum_j \delta_{ij}^k s_{ijk} \\ &= \mu + \beta_i + \tau_k + (\beta\tau)_{ik} + \omega_{ik} + \zeta_{ijk}, \end{aligned} \tag{9.53}$$

where $(\beta\tau)_{ik}$ represents the block-treatment interaction with

$$\sum_i (\beta\tau)_{ik} = \sum_k (\beta\tau)_{ik} = 0$$

and ω_{ij} and ζ_{ijk} are random variables representing unit error and strict unit-treatment interaction, respectively. We shall assume now that all $s_{ijk} = 0$ and hence model (9.53) reduces to

$$y_{ik} = \mu + \beta_i + \tau_k + (\beta\tau)_{ik} + \omega_{ik}$$

or, if we add [see model (9.23)] treatment and observational error,

$$y_{ik} = \mu + \beta_i + \tau_k + (\beta\tau)_{ik} + e_{ik}. \tag{9.54}$$

It is clear from (9.54) that $(\beta\tau)_{ik}$ and e_{ik} lead to the same entry in the ANOVA table, that is, cannot be separated, with

$$E[\text{MS}(E)] = \sigma_e^2 + \sum_{i,k} (\beta\tau)_{ik}^2 / (b-1)(t-1). \tag{9.55}$$

Since there is no mean square with expected value equal to σ_e^2 , there does not exist a test for $H_0 : (\beta\tau)_{ik} = 0$ for all i and k .

9.6.3 One Blocking Factor: A Specific Model for Nonadditivity

It is evident from (9.54) and (9.55) that the interaction as expressed by the $(b-1)(t-1)$ independent $(\beta\tau)_{ik}$'s must be modeled more specifically so that it accounts for some of the $(b-1)(t-1)$ d.f. with the remaining d.f. attributable to random error. Such a procedure was first proposed by Tukey (1949) and more explicitly by Mandel (1961). We shall describe briefly Mandel's procedure and show how it relates to Tukey's procedure.

It follows from (9.54) that

$$E(y_{ik} - \bar{y}_{i.}) = \tau_k + (\beta\tau)_{ik}$$

and under additivity, that is, all $(\beta\tau)_{ik} = 0$, that

$$E(y_{ik} - \bar{y}_{i.}) = \tau_k,$$

which is independent of i . One way to model the dependence of $E(y_{ik} - \bar{y}_{i.})$ on i is to consider a linear function of β_i and write

$$E(y_{ik} - \bar{y}_{i.}) = \tau_k + Q_k\beta_i, \quad (9.56)$$

that is, assume

$$(\beta\tau)_{ik} = Q_k\beta_i. \quad (9.57)$$

We are then considering the model

$$y_{ik} = \mu + \beta_i + \tau_k + Q_k\beta_i + e_{ik} \quad (9.58)$$

with $\sum_i \beta_i = \sum_k \tau_k = \sum_k Q_k = 0$. In order to give a concrete interpretation to model (9.58), let us write $Q_k = \gamma_k - 1$. Then (9.58) becomes

$$\begin{aligned} y_{ik} &= \mu + \tau_k + \gamma_k\beta_i + e_{ik} \\ &= \mu_k + \gamma_k\beta_i + e_{ik}, \end{aligned} \quad (9.59)$$

where $\mu_k = \mu + \tau_k$. Thus the data from a RCBD can be expressed as a set of t regression lines where the b observations for treatment k ($k = 1, 2, \dots, t$) are regressed on the block effects. It follows from $\sum_k Q_k = 0$ that

$$\bar{\gamma} = \frac{1}{t} \sum_k \gamma_k = 1. \quad (9.60)$$

If all γ_k are equal and because of (9.60) equal to 1, then (9.58) reduces to the additive model, hence departure of some γ_k from 1 indicates block-treatment interaction. Since the β_i , the regressor variables, are not known, we replace them by $\hat{\beta}_i = \bar{y}_{i.} - \bar{y}_{..}$ and hence obtain in the usual way

$$\hat{\gamma}_k = \frac{\sum_i y_{ik} \hat{\beta}_i}{\sum_i \hat{\beta}_i^2}. \quad (9.61)$$

Using $\hat{\mu} = \bar{y}_{..}$, $\hat{\tau}_k = \bar{y}_{.k} - \bar{y}_{..}$ we can then write the following identity as suggested by (9.58)

$$y_{ik} = \hat{\mu} + \hat{\beta}_i + \hat{\tau}_k + (\hat{\gamma}_k - 1)\hat{\beta}_i + \Delta_{ik} \quad (9.62)$$

with

$$\Delta_{ik} = (y_{ik} - \bar{y}_{.k}) - \hat{\gamma}_k \hat{\beta}_i.$$

9.6.4 Testing for Nonadditivity

Using the properties of the various terms in (9.62), it follows easily that

$$\sum_{i,k} y_{ik}^2 = bt\hat{\mu}^2 + t \sum_i \hat{\beta}_i^2 + b \sum_k \hat{\tau}_k^2 + \sum_k (\hat{\gamma}_k - 1)^2 \sum_i \hat{\beta}_i^2 + \sum_{ik} \Delta_{ik}^2$$

provides a partitioning of the total sum of squares and gives rise to the basic ANOVA in Table 9.5. Mandel (1961), following arguments given by Scheffé (1959), has shown that, under $H_0: \gamma_1 = \gamma_2 = \dots = \gamma_t = 1$ and the assumption of normality, SS(Slopes) follows a scaled central χ^2 -distribution with $t - 1$ d.f. Also, SS(Error) follows a scaled χ^2 -distribution with $(b - 2)(t - 1)$ d.f., the scales being the same for both sums of squares. Since both sums of squares are independently distributed it then follows that

$$F = \frac{\text{MS(Slopes)}}{\text{MS(Error)}} \quad (9.63)$$

provides a test for $H_0: \gamma_1 = \gamma_2 = \dots = \gamma_t = 1$ and hence a test for block-treatment interaction of the form specified by model (9.58). A derivation of this test based on randomization theory is provided by Roux (1984).

9.6.5 Tukey's Test for Nonadditivity

Tukey (1949) implicitly and Ward and Dick (1952) explicitly consider a special case of (9.58) by using

$$Q_k = \theta \tau_k \quad (9.64)$$

and hence the model

$$y_{ik} = \mu + \beta_i + \tau_k + \theta \beta_i \tau_k + e_{ik} \quad (9.65)$$

to detect interaction. This may seem to be a very specialized and narrow model but Ward and Dick (1952) show that (9.65) arises from a multiplicative model of the form

$$y_{ik} = (\mu' + \beta'_i + e'_{ik})(\mu'' + \tau''_k + e''_{ik})$$

with $\Sigma \beta'_i = \Sigma \tau''_k = 0$. Following earlier arguments we can write (9.64) as

$$\gamma_k - 1 = \theta \tau_k,$$

that is, the regression coefficients are expressed as a linear function of the treatment effects. Writing

$$\hat{\gamma}_k - 1 = \theta \hat{\tau}_k + \delta_k$$

Table 9.5 ANOVA for RCBD under Nonadditivity

Source	d.f.	SS
Blocks	$b - 1$	$t \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$
Treatments	$t - 1$	$b \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2$
Slopes	$t - 1$	$\sum_k (\hat{\gamma}_k - 1)^2 \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$
Regression	1	$\hat{\theta}^2 \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2$
Deviation	$t - 2$	Subtraction
Error	$(b - 2)(t - 1)$	Subtraction
Total	$bt - 1$	$\sum_{ik} (y_{ik} - \bar{y}_{..})^2$

with $\sum_k \delta_k = 0$, we obtain in the usual way the estimate of the regression coefficient θ as

$$\begin{aligned}
 \hat{\theta} &= \frac{\sum_k (\hat{\gamma}_k - 1) \hat{\tau}_k}{\sum_k \hat{\tau}_k^2} \\
 &= \frac{\sum_{ik} y_{ik} (\bar{y}_{i.} - \bar{y}_{..}) (\bar{y}_{.k} - \bar{y}_{..})}{\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2}.
 \end{aligned} \tag{9.66}$$

It follows then that in Table 9.5

$$\begin{aligned}
 \text{SS(Slopes)} &= \sum_k (\hat{\gamma}_k - 1)^2 \sum_i \hat{\beta}_i^2 \\
 &= \sum_k (\hat{\theta} \hat{\tau}_k)^2 \sum_i \hat{\beta}_i^2 + \sum_k \hat{\delta}_k^2 \sum_i \hat{\beta}_i^2 \\
 &= \hat{\theta}^2 \sum_k \hat{\tau}_k^2 \sum_i \hat{\beta}_i^2 + \sum_k \hat{\delta}_k^2 \sum_i \hat{\beta}_i^2
 \end{aligned} \tag{9.67}$$

can be partitioned into two components,

$$\text{SS(Regression)} = \hat{\theta}^2 \sum_k \hat{\tau}_k^2 \sum_i \hat{\beta}_i^2 \tag{9.68}$$

and

$$SS(\text{Deviation}) = \sum_k \hat{\delta}_k^2 \sum_i \hat{\beta}_i^2, \quad (9.69)$$

where

$$\hat{\delta}_k = \hat{\gamma}_k - \hat{\theta} \hat{\tau}_k - 1.$$

To test the hypothesis of no interaction, that is, $H_0: \theta = 0$, Tukey (1949) proposed the test statistic,

$$F = \frac{SS(\text{Regression})}{[SS(\text{Deviation}) + SS(\text{Error})]/[(b-1)(t-1)-1]}$$

or, using the notation from Table 9.2,

$$F = \frac{SS(\text{Regression})}{[SS(E) - SS(\text{Regression})]/[(b-1)(t-1)-1]}, \quad (9.70)$$

which follows an F -distribution with 1 and $(b-1)(t-1)-1$ d.f. (see Scheffé, 1959). Robinson (1975) has shown that this test is a reasonable approximation to the corresponding test based on randomization theory. The test given by (9.70) is generally referred to as Tukey's *one-degree-of-freedom test for nonadditivity*.

An alternative derivation of (9.66) and (9.70) is given by Scheffé (1959). It is based on the model

$$y_{ik} = \mu + \beta_i + \tau_k + \theta x_{ik} + e_{ik}$$

with $x_{ik} = (\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.k} - \bar{y}_{..})$, that is, replacing $\beta_i \tau_k$ in (9.65) by $\hat{\beta}_i \hat{\tau}_k$. Using the analysis of covariance technique (see Section 9.4) we obtain

$$\begin{aligned} \hat{\theta} &= \frac{E_{xy}}{E_{xx}} = \frac{\sum_{ik} (y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..})(x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x}_{..})}{\sum_{ik} (x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x}_{..})^2} \\ &= \frac{\sum_{ik} y_{ik} x_{ik}}{\sum_{ik} x_{ik}^2} \end{aligned}$$

since $\bar{x}_{i.} = \bar{x}_{.k} = \bar{x}_{..} = 0$. Hence $\hat{\theta}$ takes on the form (9.66).

9.6.6 Generalizations

A generalization of Tukey's test was proposed by Mandel (1971) and investigated in more detail by Johnson and Graybill (1972). They consider a model of the form

$$y_{ik} = \mu + \beta_i + \tau_k + \theta \alpha_i \gamma_k + e_{ik} \quad (9.71)$$

and derive a test for $H_0 : \theta = 0$ by using the $t \times b$ matrix of residuals $\mathbf{Z} = (z_{ik})$, where $z_{ik} = y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..}$. They derive the likelihood ratio test statistic

$$\Lambda^* = \left[\frac{\sum_{ik} z_{ik}^2 - \lambda_1}{\sum_{ik} z_{ik}^2} \right]^{bt/2},$$

where λ_1 is the largest eigenvalue of $\mathbf{Z}'\mathbf{Z}$. The distribution of Λ^* is then related to that of the eigenvalues of a Wishart matrix. For more details the reader is referred to Johnson and Graybill (1972), Corsten and van Eijnsbergen (1972), and Marasinghe (1985).

We close this section with the following remarks: The Johnson-Graybill procedure outlined above and the extensions due to Mandel (1971) using more than just one term to model interaction, that is, using more than just the largest eigenvalue of $\mathbf{Z}'\mathbf{Z}$, were developed for two-way layouts with one observation per cell. Even though data from a RCBD can also be presented as a two-way array, we have pointed out earlier in this chapter that there exists a certain asymmetry between blocks and treatments. For this reason we prefer to test for nonadditivity in the RCBD by using (9.63) or (9.70).

As models (9.58) and (9.65) indicate the type of interaction included in the model is of a very specific form. This limits, by necessity, our general inquiry into the possible existence of block-treatment interaction. Obviously, if the tests presented by (9.63) or (9.70) are significant then such interaction is present. If those tests, however, are not significant then this indicates only that interaction of the specific type is not present, but this does not preclude block-treatment interaction of a different type, except that we may be unable to detect it.

9.6.7 Several Blocking Factors

We now turn to the case of several blocking factors. Such cases can occur in different ways which are characterized by the different relationships of the factors with each other, that is, whether they are crossed or nested (see 4.12.2). We shall illustrate this first in terms of the following examples before we consider the question of block-treatment interaction.

EXAMPLE 9.1: An experiment was set up to study genetic parameters for radiata pine (Dean, et al., 2006). The full-sib families from a half-diallel cross with five parents represent the treatments which are evaluated for several growth traits, such as height and sectional area of stems, in a field trial using an RCBD at two different sites. More specifically, the basic arrangement consists of six blocks per site, each block containing ten plots to which the ten full-sib families are assigned at random. On each plot five trees from the assigned full-sib family are planted. Both sites are located on the North Island of New Zealand, but one site has colder climate than the other, being thus more prone to frost and needle blight. \square

EXAMPLE 9.2: The objective of an agricultural study was to determine the effect of each *G. soja* SCN resistance gene on yield and other agronomic traits in elite soybean backgrounds (Kabelka et al., 2006). One part of the field trial consists of growing plants from 100 derived genetic lines, representing the treatments, in 1.5×3.2 m two-row plots, representing the experimental units, in an RCBD with $b = 2$ blocks (replications). This basic arrangement was repeated, using different randomizations in four different environments. More specifically, the environments represent combinations of two years and different sites. Plots were evaluated for days to maturity, plant height, lodging, and seed yield. \square

EXAMPLE 9.3: A study is contemplated to examine the effectiveness of three methods to memorize German vocabulary at the high school level (Kirk, 1982). It seems reasonable to take student ability, as measured by IQ, and gender into account. Thus, one approach would be to set up, say, five IQ classes, with three students for each class and gender, leading to ten blocks. In each block the three methods under investigation will then be assigned randomly to the three students. \square

The above examples have one feature in common: They can be looked upon as *replicated randomized block experiments*, where one of the blocking factors can be considered the “replicating” factor, typically an intrinsic factor (\mathcal{Z}). In Example 9.1 the basic RCBD is replicated at different sites, indicating different climates; in Example 9.2 the replicates are the different environments, and in Example 9.3 we have an RCBD for each gender. What is different, however, is the relationship of the blocking factors to each other: In Examples 9.1 and 9.2 we have a nesting relationship, whereas in Example 9.3 we have a crossed relationship. More specifically, the blocks in the field are nested within the sites and environments, respectively, because the blocks at one site (environment) are different from the blocks at another site (environment) whereas the IQ classes are the same for both genders.

This crucial difference in the blocking factor relationship is reflected in the linear models that describe the data from such experiments as we extend model (9.23) and exploit those relationships. To show this we begin with model (9.23)

$$y_{ik} = \mu + \beta_i + \tau_k + e_{ik}, \quad (9.72)$$

where in this form the β_i represent the “block” effects of the site-block (Example 9.1) or environment-block (Example 9.2) or the IQ class-gender (Example 9.3) combinations. It is advantageous, however, to explicitly represent the individual blocking factors and their relationship to each other, because this will allow us to test certain aspects of “block”-treatment interaction which may be important for the analysis and interpretation of the experimental data.

For the general development of extending model (9.72) we shall denote the two blocking factors by A and C , with A having a “levels” and C having c “levels”. Thus, the total number of blocks is $b = ac$. We then consider the following two cases:

(i) Nested blocking factors:

For this case the β_i in (9.72) will be expanded as $\alpha_i + \gamma_{ij}$, and the new model may contain an interaction term representing possible interaction between blocking factor A and the treatments. Thus, the model may be written as

$$y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \tau_k + (\alpha\tau)_{ik} + e_{ijk} \quad (9.73)$$

with $i = 1, 2, \dots, a$; $j = 1, 2, \dots, c$; $k = 1, 2, \dots, t$; α_i representing the effect of the i -th level of A , and γ_{ij} representing the effect of the j -th level of factor C at the i -th level of factor A . The interaction term $(\alpha\tau)_{ik}$ represents part of the block-treatment interaction, namely factor A -treatment interaction ($A \times T$).

The analysis of variance associated with model (9.73) is given in Table 9.6. A derivation based on randomization theory is given by Stewart (1980) and Hinkelmann and Alcorn (1998). Table (9.6) indicates that

$$F = \frac{MS(A \times T)}{MS(E)}$$

with $(a-1)(t-1)$ and $a(c-1)(t-1)$ d.f. provides a test for part of the possible block-treatment interaction, $A \times T$.

In Example 9.1 the blocking structure is the same as in Example 9.2, but the experimental set-up provides an additional feature, namely subsampling (see Section 9.2.7) as each tree is an observational unit. An obvious extension of model (9.73) is given in (9.74), with the associated analysis of variance given in Table 9.7,

$$y_{ijkl} = \mu + \alpha_i + \gamma_{ij} + \tau_k + (\alpha\tau)_{ik} + \epsilon_{ijk} + \eta_{ijkl} \quad (9.74)$$

with $l = 1, 2, \dots, n$. It is clear from Table 9.7 that now the experimental error plays the important role in testing for $A \times T$ interaction as the denominator of the F -ratio

$$F = \frac{MS(A \times T)}{MS(EE)}$$

with $(a-1)(t-1)$ and $a(c-1)(t-1)$ d.f.

(ii) Crossed blocking factors:

Each combination of the levels of the factors A and C represents a block. This factorial structure (not to be confused with factorial treatment structure; see Section 11.2) leads to an expansion of β_i in (9.72) into $\alpha_i + \gamma_j + (\alpha\gamma)_{ij}$ and to inclusion of certain block-treatment interaction terms as given in model (9.75):

$$y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \tau_k + (\alpha\tau)_{ik} + (\gamma\tau)_{jk} + e_{ijk} \quad (9.75)$$

with $i = 1, 2, \dots, a$; $j = 1, 2, \dots, c$; $k = 1, 2, \dots, t$. The $\alpha_i, \gamma_j, (\alpha\gamma)_{ij}$ represent block effect components, and $(\alpha\gamma)_{ik}, (\gamma\tau)_{jk}$ represent interactions between the blocking factors A, C and the treatments. Model (9.75) leads to an analysis of variance as given in Table 9.8.

Table 9.6 ANOVA for RCBD with Nested Blocking Factors

Source	d.f.	SS	E(MS)
Blocks	$ac - 1$	$t \sum_{ij} (\bar{y}_{ij.} - \bar{y}_{...})^2$	—
A	$a - 1$	$tc \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	—
$C(A)$	$a(c - 1)$	$t \sum_{ij.} (\bar{y}_{ij.} - \bar{y}_{i..})^2$	—
Treatments (T)	$t - 1$	$ac \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$	$\sigma_e^2 + \frac{ac \sum_k \tau_k^2}{t - 1}$
$A \times T$	$(a - 1)(t - 1)$	$c \sum_{ik} (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$	$\sigma_e^2 + \frac{c \sum_{ik} (\alpha\gamma)_{ik}^2}{t - 1}$
Error	$a(c - 1)(t - 1)$	$\sum_{ijk} (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} + \bar{y}_{i..})^2$	σ_e^2
Total	$act - 1$	$\sum_{ijk} (y_{ijk} - \bar{y}_{...})^2$	

Table 9.7 ANOVA for RCBD with Nested Blocking Factors and Subsampling

Source	d.f.	SS	E(MS)
A	$a - 1$	$tcn \sum_i (\bar{y}_{i...} - \bar{y}_{....})^2$	—
$C(A)$	$a(c - 1)$	$tn \sum_{ij} (\bar{y}_{ij..} - \bar{y}_{i...})^2$	—
T	$t - 1$	$acn \sum_k (\bar{y}_{..k.} - \bar{y}_{....})^2$	$\sigma_\eta^2 + n\sigma_\epsilon^2 + \frac{cn}{t-1} \sum_k \tau_k^2$
$A \times T$	$(a - 1)(t - 1)$	$cn \sum_{ik} (\bar{y}_{i.k.} - \bar{y}_{i...} - \bar{y}_{..k.} + \bar{y}_{....})^2$	$\sigma_\eta^2 + n\sigma_\epsilon^2 + \frac{cn}{(a-1)(t-1)} \sum_{ik} (\alpha\gamma)_{ik}^2$
Exp. Error	$a(c - 1)(t - 1)$	$n \sum_{ijk} (\bar{y}_{ijk.} - \bar{y}_{ij..} - \bar{y}_{i.k.} + \bar{y}_{i...})^2$	$\sigma_\eta^2 + n\sigma_\epsilon^2$
Obs. Error	$act(n - 1)$	$\sum_{ijkl} (\bar{y}_{ijkl} - \bar{y}_{ijk.})^2$	σ_η^2
Total	$actn - 1$	$\sum_{ijkl} (\bar{y}_{ijkl} - \bar{y}_{....})^2$	

Table 9.8 ANOVA for RCBD with Crossed Blocking Factors

Source	d.f.	SS	E(MS)
Blocks	$ac - 1$	$t \sum_{ij} (\bar{y}_{ij.} - \bar{y}_{...})^2$	—
A	$a - 1$	$ct \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	—
C	$c - 1$	$at \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$	—
$A \times C$	$(a - 1)(c - 1)$	$t \sum_{ij} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	—
Treatments (T)	$t - 1$	$ac \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$	$\sigma_e^2 + \frac{ac \sum_k \tau_k^2}{t - 1}$
$A \times T$	$(a - 1)(t - 1)$	$c \sum_{ik} (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$	$\frac{c \sum_{ik} (\alpha\tau)_{ik}^2}{\tau_{ik}(a - 1)(t - 1)}$
$C \times T$	$(c - 1)(t - 1)$	$a \sum_{jk} (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$	$\sigma_e^2 + \frac{a \sum_{jk} (\gamma\tau)_{jk}^2}{\tau_{jk}(c - 1)(t - 1)}$
Error	$(a - 1)(c - 1)(t - 1)$	$\sum_{ijk} (\bar{y}_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...})^2$	σ_e^2
Total	$act - 1$	$\sum_{ijk} (\bar{y}_{ijk} - \bar{y}_{...})^2$	

In this situation it follows from Table 9.8 that we can isolate and test for two block-treatment interaction components, namely $A \times T$ and $C \times T$ by using the F -ratios

$$F = \frac{MS(A \times T)}{MS(E)}$$

with $(a - 1)(t - 1)$ and $(a - 1)(c - 1)(t - 1)$ d.f. and

$$F = \frac{MS(C \times T)}{MS(E)}$$

with $(c - 1)(t - 1)$ and $(a - 1)(c - 1)(t - 1)$ d.f., respectively.

9.6.8 Dealing with Block-Treatment Interaction

In the preceding sections we have explored methods of investigating, that is, detecting possible block-treatment interactions for various types of randomized complete block designs. If interactions do, indeed, exist then the question arises: What to do next? How can we or, even, can we at all make inferences about the treatment effects? There do not seem to exist general answers to these questions.

What are the problems? The investigator wants to compare and make recommendations about treatments. But the existence of block-treatment interaction implies that such comparisons are not the same for all blocks. Hence making comparisons in the usual way, that is, by comparing treatment means, may present a wrong picture. Also, as Kempthorne (1952, Section 8.3) showed, with nonadditivity it is not possible to attach “reasonable” standard errors to treatment comparisons. And finally, nonadditivity in a two-way table may be due to interaction or to nonhomogeneous variances as pointed out, for example, by Snee (1982) who also showed how knowledge of the subject matter can be important in explaining nonadditivity. And modeling such nonadditivity may prove to be an important aspect of data analysis.

It is here that the distinction between RCBDs with one blocking factor and two (or more) blocking factors becomes important. And beyond that, for the one blocking factor situation we distinguish between whether the blocking factor is a nonspecific (\mathcal{U}) or an intrinsic (\mathcal{Z}) factor.

The reason for making this distinction is that nonadditivity for these two types of RCBDs may lead to different actions. Clearly, for the first type any attempt of explaining or modeling nonadditivity is of no value with regard to comparing treatments. Rather, it may be helpful to remove such nonadditivity through a suitable transformation using methods described in Section 6.10. In this case it may be useful to plot residuals $(y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..})$ against the observations y_{ik} to gain some insight into the form of nonadditivity and hence obtain an idea what type of transformation may be appropriate.

With regard to RCBDs of the second type it may indeed be important to model possible nonadditivity as a means of interpreting differential treatment effects. In fact, in this case block-treatment interactions may be more important than treatment effects themselves. We argue that then, if at all possible, a different design should have been used, namely a generalized randomized block design as discussed in the next section.

The situation is somewhat clearer and more options may be available if we have more than one blocking factor. Based on the analysis of variance for models (9.73), (9.74) and (9.75) we have displayed F -ratios that can be used to test whether certain components of block-treatment interaction, namely that between intrinsic blocking factor and treatments, for example, $A \times T$ and/or $C \times T$, are present.

To discuss this in more detail let us first consider model (9.73) in the context of Example 9.1.

EXAMPLE 9.1 (continued): The type of question that is often asked is: Are there differences between the two sites? We point out that not only is this the wrong question, but also that such a hypothesis cannot be tested since the factor “site” is a blocking factor (see Section 9.2.6). Rather, what is really meant is: Are the treatments (all or some) performing differently at the two sites? And this is to ask the question whether there exists site-treatment interaction, that is, $A \times T$ interaction.

If the $A \times T$ interaction is significant further investigation is required to enunciate to what extent and in which way the full-sib families, or generally, the treatments, perform differently at the two sites. This can be done in what is referred to as an *interaction plot* by plotting the treatment means separately for each site (see also Section 9.7.4). The two graphs are not essentially parallel (as they would be if interaction is not present), but can take on various forms. If they move essentially in the same direction then we have what we call *codirectional interaction* (see Section 9.7.4) or *synergistic interaction* (see van Belle et al., 2004). In this case it is still possible and informative to consider inference about the overall treatment effects either in the form of tests of hypotheses, including various types of comparisons, or confidence interval estimation. If the interaction is not codirectional then it will be more appropriate to consider inference separately for each site. This can be done either in the context of model (9.73), that is, by using the error term from the ANOVA of Table 9.6 or by analyzing the data from each site as separate RCBDs.

Finally, if the $A \times T$ interaction is not significant then we may contemplate to delete the interaction term $(\alpha\tau)_{ik}$ from model (9.73) and reanalyze the data with the new model. This is largely a philosophical issue, and we take the point of view that if deleting the term at all it should be handled in the context of a preliminary test (Bancroft, 1964) and be dropped only if $P > .25$, say. The effect of this will be, of course, an increase in the d.f. for error from $a(c-1)(t-1)$ to $(ac-1)(t-1)$. \square

We now turn to Example 9.3 and model (9.75).

EXAMPLE 9.3 (continued): Much of what we have said for Example 9.1 holds also here. Instead of only one interaction term we now deal with two interaction terms and hence potentially two interaction plots, one for $A \times T$ and one for $C \times T$. If both interactions are significant and not codirectional then we face the same problem discussed earlier for the case of one intrinsic blocking factor. Here, again, a generalized randomized block design may then be a better option to repeat this experiment.

If one and/or the other interaction is not significant then we may contemplate pooling the non-significant interaction(s) with the error term in the ANOVA of Table 9.8.

\square

To summarize our discussion, we emphasize that it is important not to ignore the

possibility of block-treatment interaction, especially when intrinsic blocking factors are involved. In that case the interaction may be more important, and the subject-matter specialist may be able to provide important input and insight. And this should be reflected in the experimental design and the subsequent analysis. Many considerations come into play here, and it is impossible to provide specific directions for all cases.

9.7 GENERALIZED RANDOMIZED BLOCK DESIGN

9.7.1 Definition

As mentioned in Sections 9.6.7 and 9.6.8 there exist situations where block-treatment interaction is strongly suspected a priori and where such interaction may be the major focus of the investigation and hence explicit characterization of its form is of utmost importance. This may occur, for example, when an intrinsic blocking factor is introduced by choice to broaden the inference from the experiment, for instance, different varieties of plants. The *generalized randomized block design* (GRBD) to be discussed in this section is the most appropriate design for such situations.

We call a block design a GRBD if we have b blocks, each block containing $s = rt$ EUs, such that each of the t treatments is applied to r EUs in each block (note that for $r = 1$ we have, of course, the RCBD). The treatments are assigned randomly to the EUs, and independent randomizations are used for different blocks.

9.7.2 Derived Linear Model

Let T_{ijk} denote the conceptual response if treatment k is applied to the j th EU in the i th block. We can then write the following identity:

$$\begin{aligned} T_{ijk} = & \bar{T}_{...} + (\bar{T}_{i..} - \bar{T}_{...}) + (\bar{T}_{ij.} - \bar{T}_{i..}) + (\bar{T}_{..k} - \bar{T}_{...}) \\ & + (\bar{T}_{i.k} - \bar{T}_{i..} - \bar{T}_{..k} + \bar{T}_{...}) + (T_{ijk} - \bar{T}_{ij.} - \bar{T}_{i.k} + \bar{T}_{i..}). \end{aligned} \quad (9.76)$$

The physical interpretation of most of the terms in (9.76) have been given in Section 9.2. In addition we now have the term

$$(\beta\tau)_{ik} = (\bar{T}_{i.k} - \bar{T}_{i..}) - (\bar{T}_{..k} - \bar{T}_{...})$$

the difference between the effect of treatment k in block i and the overall effect of treatment k . To the extent that this term is different from zero, this is a measure of block-treatment interaction. Also, the term

$$(T_{ijk} - \bar{T}_{ij.}) - (\bar{T}_{i.k} - \bar{T}_{i..}),$$

that is, the difference between the effect of treatment k on EU j in block i and the effect of treatment k in block i , is a measure of the unit-treatment interaction. We shall henceforth assume that such interaction is negligible. We can then write (9.76) as

$$T_{ijk} = \mu + \beta_i + u_{ij} + \tau_k + (\beta\tau)_{ik} \quad (9.77)$$

with

$$\begin{aligned}\sum_i \beta_i &= 0, & \sum_k \tau_k &= 0 \\ \sum_{j=1}^s u_{ij} &= 0 & \text{for every } i \\ \sum_i (\beta\tau)_{ik} &= \sum_k (\beta\tau)_{ik} = 0.\end{aligned}$$

The actual experiment then consists of randomly assigning each treatment to r EUs in each block using independent randomizations in different blocks. The procedure, using SAS PROC PLAN, is illustrated in Table 9.9, where in the output the treatment numbers are superimposed on the unit numbers within each block.

The randomization process is characterized by the design random variables

$$\delta_{ij}^k = \begin{cases} 1 & \text{if treatment } k \text{ is applied to the } j \text{ th EU in block } i \\ 0 & \text{otherwise.} \end{cases}$$

Let y_{ikl} denote the observation for the l th replication of treatment k in block i ($i = 1, 2, \dots, b; k = 1, 2, \dots, t; l = 1, 2, \dots, r$). We then have

$$P(\delta_{ij}^k = 1) = \frac{r}{s} = \frac{1}{t}$$

and hence

$$E_R(\delta_{ij}^k) = \frac{1}{t}$$

and other properties of the δ_{ij}^k can be established easily. The connection between the conceptual response T_{ijk} and the actually observed response y_{ikl} is then given by

$$\sum_{l=1}^r y_{ikl} = \sum_{j=1}^s \delta_{ij}^k T_{ijk},$$

or

$$\sum_l y_{ikl} = r[\mu + \beta_i + \tau_k + (\beta\tau)_{ik}] + \sum_j \delta_{ij}^k u_{ij}. \quad (9.78)$$

This is a model based on randomization only and does not include technical errors. If we add technical errors the model (9.78) becomes (see Section 9.2)

$$\sum_l y_{ikl} = r[\mu + \beta_i + \tau_k + (\beta\tau)_{ik}] + \sum_j \delta_{ij}^k u_{ij} + \sum_l \nu_{ikl} + \sum_l \eta_{ikl} \quad (9.79)$$

with the usual assumptions for the treatment errors, ν_{ikl} , and observational errors, η_{ikl} .

9.7.3 The ANOVA Table

Using the identity

$$y_{ikl} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.k.} - \bar{y}_{...}) + (\bar{y}_{ik.} - \bar{y}_{i..} - \bar{y}_{.k.} + \bar{y}_{...}) + (y_{ikl} - \bar{y}_{ik.})$$

and writing, based on (9.79),

$$y_{ikl} = \mu + \beta_i + \tau_k + (\beta\tau)_{ik} + e_{ikl} \quad (9.80)$$

we obtain the ANOVA as given in Table 9.10. The expected values of the mean squares as given in Table 9.10 can be obtained by using the distributional properties of the δ_{ij}^k s (see Wilk, 1955) and of the ν_{ikl} and the η_{ikl} . Here we have defined

$$\sigma_u^2 = \frac{1}{b(s-1)} \sum_{ij} u_{ij}^2$$

$$\sigma_\nu^2 = E(\nu_{ikl}^2)$$

$$\sigma_\eta^2 = E(\eta_{ikl}^2).$$

We remind the reader that $\sigma_u^2 + \sigma_\nu^2 = \sigma_\varepsilon^2$ is the experimental error variance, σ_η^2 is the observational error variance, and $\sigma_\varepsilon^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$ is the overall error variance. We note also that the GRBD is an unbiased design in Yates' sense, just as the RCBD, in that under $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$ we have $E[\text{MS}(T)] = E[\text{MS}(E)]$. Further, again just as in the RCBD, the ANOVA does not lend itself to a legitimate test for block effects because $E[\text{MS}(B)] \neq E[\text{MS}(E)]$ for $\beta_1 = \beta_2 = \dots = \beta_b = 0$.

The form of the $E(\text{MS})$ in Table 9.10 suggests the following tests:

- (i) The test statistic for no block-treatment interaction, that is, $H_0 : (\beta\tau)_{ik} = 0$ for every i, k is given by

$$F = \frac{\text{MS}(B \times T)}{\text{MS}(E)} \quad (9.81)$$

with $(b-1)(t-1)$ and $bt(r-1)$ d.f.

- (ii) The test statistic for no treatment differences, that is, $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$ is given by

$$F = \frac{\text{MS}(T)}{\text{MS}(E)} \quad (9.82)$$

with $t-1$ and $bt(r-1)$ d.f., but see Section 9.7.4

Wilk (1955) has shown that these F -tests are reasonably good approximations to the corresponding randomization tests.

Table 9.10 ANOVA for GRBD

Source	d.f.	SS	$E(\text{MS})$
Blocks	$b - 1$	$tr \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \text{SS}(B)$	$\sigma_\eta^2 + \sigma_\nu^2 + tr \sum_i \beta_i^2 / (b - 1)$
Treatments	$t - 1$	$br \sum_k (\bar{y}_{.k.} - \bar{y}_{...})^2 = \text{SS}(T)$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2 + br \sum_k \tau_k^2 / (t - 1)$
Block \times Treatment Interaction	$(b - 1)(t - 1)$	$r \sum_{ik} (\bar{y}_{ik.} - \bar{y}_{i..} - \bar{y}_{.k.} + \bar{y}_{...})^2 = \text{SS}(B \times T)$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2 + r \sum_{ik} (\beta\tau)_{ik}^2 / (b - 1)(t - 1)$
Error	$btr - 1$	$\sum_{ikl} (y_{ikl} - \bar{y}_{ik.})^2 = \text{SS}(E)$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2$
Total	$btr - 1$	$\sum_{ikl} (y_{ikl} - \bar{y}_{...})^2$	

9.7.4 Analyzing Block-Treatment Interaction

We shall comment briefly on the usefulness and appropriateness of testing $H_0 : \tau_1 = \dots = \tau_t = 0$ when the test for block-treatment interaction has been found significant. In such a case it is important to study, mainly by plotting, the nature of the interaction. Following the discussion in Section 9.6.8 it is useful to plot $\bar{y}_{ik.}$ versus $\bar{y}_{i..}$ (or versus i) for each k . Alternatively, we may plot \bar{y}_{ikl} versus $\bar{y}_{.k.}$ (or versus k) for each i (see Section 9.6.8). Two examples are given in Figures 9.1 and 9.2 based on the (fictitious) data in Table 9.11.

The important feature of Figure 9.1a and 9.2a is that the changes from block to block, though different for the various treatments, are in the same direction. We have referred to this type of interaction as *codirectional interaction*. If we define

$$\tau_{ik} = \bar{T}_{i.k} - \bar{T}_{i..}$$

as the effect of treatment k in block i and

$$\hat{\tau}_{ik} = \bar{y}_{ik.} - \bar{y}_{i..}$$

as its estimator ($i = 1, 2, \dots, b$; $k = 1, 2, \dots, t$) then codirectional interaction implies that if $\tau_{ik} - \tau_{i'k} \geq 0$, then also $\tau_{ik'} - \tau_{i'k'} \geq 0$ and $(\tau_{ik} - \tau_{i'k}) \neq (\tau_{ik'} - \tau_{i'k'})$ with similar statements for the $\hat{\tau}_{ik}$. Since the “trend” for each treatment is in the same direction, it may make sense and, in fact, it may be quite useful to compare “the” effects, that is, the

$$\tau_k = \frac{1}{b} \sum_i \tau_{ik}$$

or, correspondingly, their estimators

$$\hat{\tau}_k = \frac{1}{b} \sum_i \hat{\tau}_{ik} = \bar{y}_{.k.} - \bar{y}_{...} \quad (9.83)$$

Suppose, for example, that the blocks represent breeds of cattle and the treatments represent feeding regimens. In order to avoid unnecessary complications let us assume that the data given in Table 9.11a actually are the $\bar{T}_{i.k}$, representing, for example, units of weight gain per week. Although the difference in gain between regimens 1 and 2 is small for breed 1, zero for breed 2, and quite substantial for breeds 3 and 4, it seems to be useful information that on average the gain due to regimen 2 is 2.5 units higher than that from regimen 1. This is the gain a farmer would realize using regimen 2 if he had (equal proportions of) cattle from all four breeds.

The picture is obviously less clear if the outcome of the experiment is represented by the data in Table 9.11b and presented in Figures 9.1b and 9.2b. Even though in this case regimen 3 is the best on average, it is clear that regimen 1 is best for breeds 3 and 4 and regimen 3 is best for breeds 1 and 2. This form of what we might call *antidirectional interaction* or *antagonistic interaction* (see van Belle et al., 2004), manifested by the fact that not only are the differences $\tau_{ik} - \tau_{i'k}$ different for different k , but they also differ possibly in sign (direction), obviously dictates different action for the different breeds and hence makes consideration of the overall regimen effects meaningless.

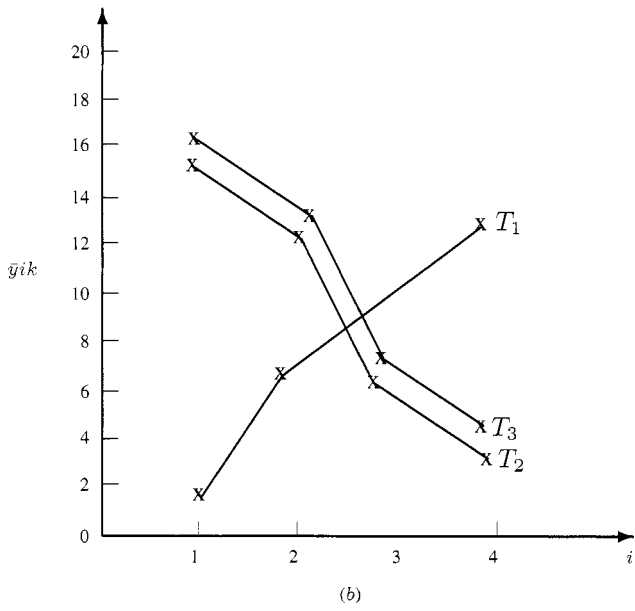
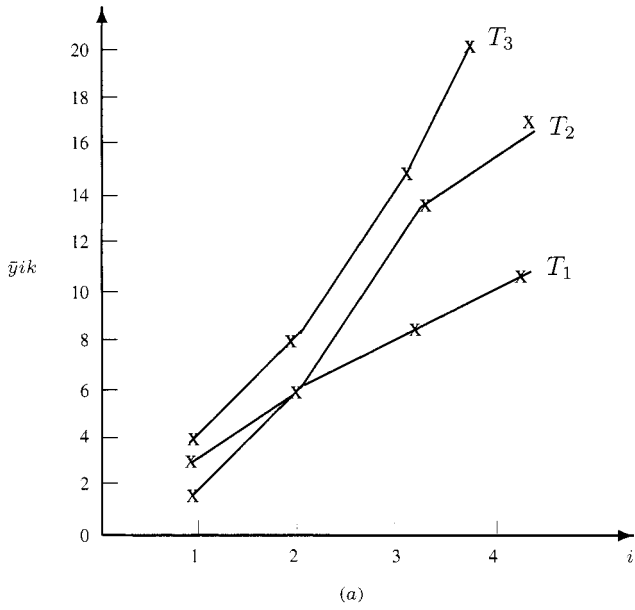


Figure 9.1 Interaction plots (\bar{y}_{ik} vs i)

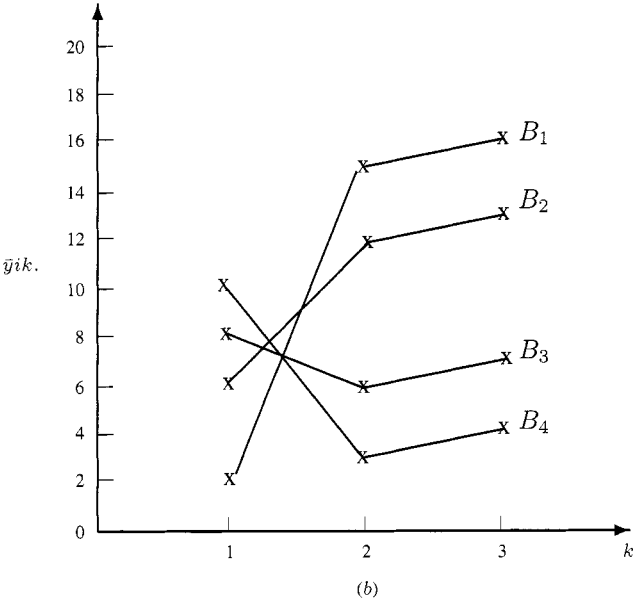
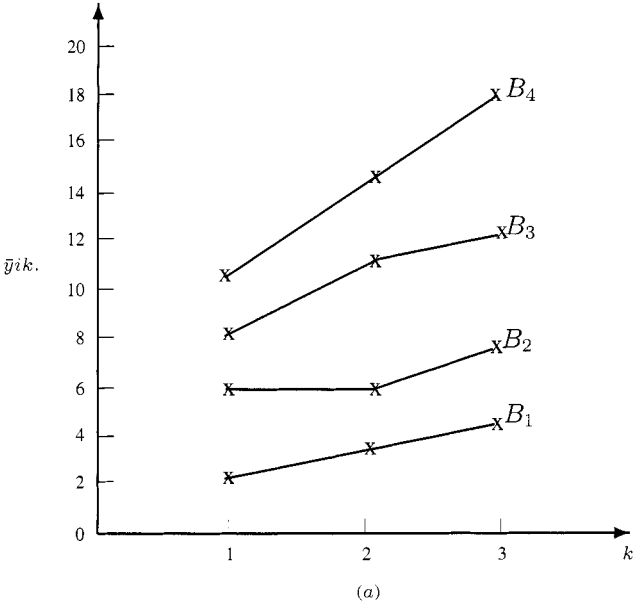


Figure 9.2 Interaction plots (\bar{y}_{ik} vs k)

Table 9.11 Block-Treatment Averages (\bar{y}_{ik}) for GRBD

(a)		Block				$\bar{y}_{.k}$
		1	2	3	4	
Treatment	1	2	6	8	10	6.50
	2	3	6	12	15	9.00
	3	4	8	13	20	11.25
	$\bar{y}_{i..}$	3	6.67	11.00	15.00	

(b)		Block				$\bar{y}_{.k}$
		1	2	3	4	
Treatment	1	2	6	8	10	6.50
	2	15	12	6	3	9.00
	3	16	13	7	4	10.00
	$\bar{y}_{i..}$	11.00	10.33	7.00	5.67	

Comparisons of the τ_{ik} 's within block i are, of course, estimated by

$$\sum_k c_{ik} \hat{\tau}_{ik} = \sum_k c_{ik} \bar{y}_{ik.} \left(\sum_k c_{ik} = 0 \right) \quad (9.84)$$

with

$$\text{var} \left(\sum_k c_{ik} \hat{\tau}_{ik} \right) = \sum_k c_{ik}^2 \frac{\sigma_e^2}{r}$$

and

$$\hat{\text{var}} \left(\sum_k c_{ik} \hat{\tau}_{ik} \right) = \sum_k c_{ik}^2 \frac{\hat{\sigma}_e^2}{r} \quad (9.85)$$

with

$$\hat{\sigma}_e^2 = \text{MS}(E).$$

Using $\hat{\sigma}_e^2 = \text{MS}(E)$ from Table 9.10 shows that, even though we consider treatment comparisons (9.84) within blocks, we are using the overall (pooled) error variance estimate for purposes of inference. An alternative procedure is to consider the observations in each block as the outcome of a CRD, say $\text{CRD}_1, \text{CRD}_2, \dots, \text{CRD}_b$, and

analyze them accordingly. This means that rather than estimating the variance of an estimated contrast by (9.85) we would use

$$\text{var} \left(\sum_k c_{ik} \hat{\tau}_{ik} \right) = \sum_k c_{ik}^2 \hat{\sigma}_{e(i)}^2 \quad (9.86)$$

for $i = 1, 2, \dots, b$, where $\hat{\sigma}_{e(i)}^2$ is the error mean square from the ANOVA of CRD_{*i*} with $t(r-1)$ d.f. rather than $bt(r-1)$ d.f. as in (9.85). If the number of d.f. becomes an issue, that is, if $t(r-1)$ is rather small, we may consider pooling the variance estimators $\hat{\sigma}_{e(i)}^2$ ($i = 1, 2, \dots, b$) by using a preliminary test for

$$H_0 : \sigma_{e(1)}^2 = \sigma_{e(2)}^2 = \dots = \sigma_{e(b)}^2, \quad (9.87)$$

for example, the F -max test of Hartley (1950). If (9.87) is not rejected at, say, $\alpha = .25$ then we may choose to pool and hence return to (9.85).

If meaningful, comparisons among the overall treatment effects τ_k are estimated by

$$\sum_k c_k \hat{\tau}_k = \sum_k c_k \bar{y}_{..k} \quad \left(\sum_k c_k = 0 \right)$$

with

$$\text{var} \left(\sum_k c_k \hat{\tau}_k \right) = \sum_k c_k^2 \frac{\sigma_e^2}{br}$$

and

$$\text{var} \left(\sum_k c_k \hat{\tau}_k \right) = \sum_k c_k^2 \frac{\text{MS}(E)}{br}. \quad (9.88)$$

9.7.5 A More General Formulation

The discussion so far assumes what is usually referred to as a *fixed effects model*. By that we mean that the blocks used in the experiment are the only blocks available, that each block consisted of exactly tr EUs and that the treatments used are the only treatments of interest to the investigator. Such assumptions imply the way tests of hypotheses are performed as described above. It also defines the reference population to which the results of the experiment apply. For example, the assertion of treatment differences can, strictly speaking, only be made with respect to the blocks and the EUs that were part of the experiment. A somewhat wider reference population, however, can be considered and used to derive an appropriate model for the observations from a GRBD and also to derive appropriate statistical tests. Following Wilk and Kempthorne (1955, 1956) and Zyskind (1962) we shall describe briefly such a situation without providing all the details.

Suppose we have a population of B blocks, each block containing S EUs, and a population of T treatments. The experiment then consists of selecting at random b blocks, s EUs in each block, and t treatments. The selected treatments are then randomly assigned to the EUs in each block such that each treatment occurs r times in each

Table 9.12 General Forms of $E(\text{MS})$ for GRBD

Source	d.f.	$E(\text{MS})$
Blocks	$b - 1$	$\sigma_\eta^2 + \sigma_\nu^2 + \frac{S-s}{S}\sigma_u^2 + \frac{T-t}{T}r\sigma_{\beta\tau}^2 + tr\sigma_\beta^2$
Treatments	$t - 1$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2 + \frac{B-b}{B}r\sigma_{\beta\tau}^2 + rb\sigma_\tau^2$
Block-Treatment Interaction	$(b-1)(t-1)$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2 + r\sigma_{\beta\tau}^2$
Error	$bt(r-1)$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2$
Total	$btr - 1$	

block, hence $s = tr$. A derived linear model can be obtained by introducing sampling and design random variables to link the conceptual responses and linear functions of them to the observed responses much in the same way as we have described the general idea earlier. Such a model leads to the same partitioning of the total sum of squares as given in Table 9.10, but it leads to different $E(\text{MS})$. The $E(\text{MS})$, following Wilk and Kempthorne (1956) are given in Table 9.12.

Here σ_ν^2 and σ_η^2 are defined as before and

$$\begin{aligned}\sigma_u^2 &= \frac{1}{B(S-1)} \sum_{i=1}^B \sum_{j=1}^S u_{ij}^2 \\ \sigma_{\beta\tau}^2 &= \frac{1}{(B-1)(T-1)} \sum_{i=1}^B \sum_{k=1}^T (\beta\tau)_{ik}^2 \\ \sigma_\beta^2 &= \frac{1}{B-1} \sum_{i=1}^B \beta_i^2 \\ \sigma_\tau^2 &= \frac{1}{T-1} \sum_{k=1}^T \tau_k^2\end{aligned}$$

with the u_{ij} , $(\beta\tau)_{ik}$, β_i , τ_k defined as before except now in the context of the populations from which we sample.

9.7.6 Random Block Effects

We note that for the case $b = B$, $s = S$, $t = T$ the $E(\text{MS})$ in Table 9.12 reduce to those of Table 9.10. Another extreme and important case is that where B is much larger than b (which we denote by $b \ll B$), $s = S$, $t = T$. Then $(B-b)/B \approx 1$, and the form of

the $E(\text{MS})$ in Table 9.12 suggests to test the hypothesis $H_0 : \tau_1 = \tau_2 = \dots = \tau_t$ by using the F -test

$$F = \frac{\text{MS}(T)}{\text{MS}(B \times T)} \quad (9.89)$$

with $t - 1$ and $(b - 1)(t - 1)$ d.f., which is, of course, different from (9.82). This situation is referred to as a *mixed model* (see Section 4.18) situation with block effects as random effects and treatment effects as fixed effects. It follows then also that

$$\text{var} \left(\sum_k c_k \hat{\tau}_k \right) = \sum_k c_k^2 (\sigma_e^2 + r \sigma_{\beta\tau}^2) / br$$

and

$$\hat{\text{var}} \left(\sum_k c_k \hat{\tau}_k \right) = \sum_k c_k^2 \text{MS}(B \times T) / br, \quad (9.90)$$

where $\sum_k c_k = 0$.

The situation $t \ll T$, though not inconceivable, rarely occurs in the context of comparative experiments. Hence we shall neither discuss the mixed model with block effects fixed and treatment effects random, nor the random effects model with $b \ll B$ and $t \ll T$.

We conclude this section with the following comments:

- (i) In practical situations it is sometimes not easy to decide whether block effects should be treated as random effects. For example, are the blocks in a field experiment randomly selected from a larger population of blocks? Most likely they were the only blocks available for the experiment. Or, if the experiment is replicated over two years (setting up a GRBD with nested blocking factors), are those years randomly selected? Certainly not, but still the researcher may want to consider them as “random” years. But if one year turns out to be a dry year and the other to be a wet year, then clearly we have an intrinsic blocking factor with fixed effects. There are obviously many variations of this discussion and thus this question becomes rather philosophical and often controversial. Generally, we prefer to consider the block effects as fixed effects.
- (ii) If the block effects are considered to be random effects, then this will affect the properties of the treatment least squares means, $\bar{y}_{.k.}$. We find from model (9.80) that, since $E(\beta_i) = 0$ and $E[(\beta\tau)_{ik}] = 0$,

$$E(\bar{y}_{.k.}) = \mu + \tau_k \quad (9.91)$$

and

$$\text{var}(\bar{y}_{.k.}) = \left(\sigma_{\beta}^2 + \sigma_{\beta\tau}^2 + \frac{\sigma_e^2}{r} \right) / b \quad (9.92)$$

but, for $\sum c_k = 0$,

$$\text{var} \left(\sum_k c_k \bar{y}_{.k.} \right) = \sum_k c_k^2 \left(\frac{\sigma_{\beta\tau}^2}{b} + \frac{\sigma_e^2}{br} \right) \quad (9.93)$$

or

$$\widehat{\text{var}} \left(\sum_k c_k \bar{y}_{.k} \right) = \sum_k c_k^2 \frac{\text{MS}(B \times T)}{br}$$

(see (9.90)). We note that it follows from (9.91), (9.92) and (9.93) that also for the RCBD with random block effects only $\text{var}(\bar{y}_{.k})$ will be affected, that is, the variance of a treatment least squares mean will be larger than the corresponding variance for the fixed effects case (see Example 9.16 in Section 9.10).

9.7.7 Using Satterthwaite's Procedure

In many situations the “truth” lies between the two extremes; that is, the effects are neither “fixed” nor “random” in the Eisenhart (1947) sense. The $E(\text{MS})$ in Table 9.12 give then some idea what the proper “error” term should be for testing hypotheses. For example, if $b < B$ (but $B < \infty$), $t = T$, in order to test $H_0 : \tau_1 = \tau_2 = \dots = \tau_t = 0$, that is, $H_0 : \sigma_\tau^2 = 0$, we construct a synthetic mean square

$$\text{MS}(R) = \phi_1 \text{MS}(E) + \phi_2 \text{MS}(B \times T) \quad (9.94)$$

such that

$$E[\text{MS}(R)] = E[\text{MS}(T)] - rb\sigma_\tau^2$$

that is,

$$E[\text{MS}(R)] = E[\text{MS}(T) | \sigma_\tau^2 = 0].$$

From Table 9.12 we infer immediately that

$$\phi_1 + \phi_2 = 1$$

and

$$\phi_2 = \frac{B-b}{B}.$$

Hence, (9.94) becomes

$$\text{MS}(R) = \frac{b}{B} \text{MS}(E) + \frac{B-b}{B} \text{MS}(B \times T). \quad (9.95)$$

To test H_0 we then use a procedure due to Satterthwaite (1946) which in general can be described as follows: Suppose we have random variables $X_i (i = 1, 2, \dots, m)$ which are independently distributed as $\mu_i \chi_{\nu_i}^2 / \nu_i$ where $\mu_i = E(X_i)$ and ν_i is the number of d.f. of $X_i (i = 1, 2, \dots, m)$. We then consider a random variable X defined as

$$X = \sum_{i=1}^m \phi_i X_i, \quad (9.96)$$

that is, a linear combination of independent χ^2 -distributed random variables, with

$$E(X) = \mu = \sum_{i=1}^m \phi_i \mu_i$$

and approximate X by a random variable of the form $\mu\chi^2_\nu/\nu$ where ν is determined such that X and $\mu\chi^2_\nu/\nu$ have the same variance. It follows from (9.96) then that

$$2\mu^2/\nu = 2 \sum_{i=1}^m \phi_i^2 \mu_i^2 / \nu_i$$

or

$$\nu = \mu^2 / \left(\sum_i \phi_i^2 \mu_i^2 / \nu_i \right).$$

In our case we have $m = 2$ and $X_1 = \text{MS}(E)$, $X_2 = \text{MS}(B \times T)$, $X = \text{MS}(R)$, $\phi_1 = b/B$, $\phi_2 = (B-b)/B$, $\nu_1 = bt(r-1)$, $\nu_2 = (b-1)(t-1)$. Since $\mu_1 = E[\text{MS}(E)]$ and $\mu_2 = E[\text{MS}(B \times T)]$ are unknown, we approximate ν by

$$\hat{\nu} = X^2 / \left(\sum_i \phi_i^2 X_i^2 / \nu_i \right), \quad (9.97)$$

that is,

$$\hat{\nu} = \frac{\left[\frac{b}{B} \text{MS}(E) + \frac{B-b}{B} \text{MS}(B \times T) \right]^2}{\frac{\left[\frac{b}{B} \text{MS}(E) \right]^2}{bt(r-1)} + \frac{\left[\frac{B-b}{B} \text{MS}(B \times T) \right]^2}{(b-1)(t-1)}}.$$

The test statistic for an approximate F -test for testing H_0 is then given by

$$F = \frac{\text{MS}(T)}{\text{MS}(R)} \quad (9.98)$$

with $(t-1)$ and $\hat{\nu}$ d.f.

The synthetic "error" mean square, $\text{MS}(R)$, in (9.98) and as defined in (9.95) is also used in approximate tests concerning individual treatment contrasts $\sum c_k \tau_k$ with $\sum c_k = 0$. Specifically, we may use

$$F = \frac{\left(\sum_k c_k \hat{\tau}_k \right)^2 / \left(\sum_k c_k^2 / br \right)}{\text{MS}(R)}$$

with 1 and $\hat{\nu}$ [see (9.97)] d.f. as the test statistic for testing $H_0 : \sum_k c_k \tau_k = 0$.

For the overall F -test as well as for the test concerning individual contrasts the inference space is that of the treatments used in the experiment and the population of blocks available for the experiment. Probability statements associated with the tests are valid only over this space. One should keep this in mind when extrapolating results to entities not in the population. This situation does indeed occur quite often when we want to apply the results of the experiment to entities that will arise in the future. For example, if the blocks are litters of mice then the results do not automatically carry over to litters to be born in the future. We may feel more comfortable with

our extrapolation if the future litters bear some relationship to the current litters, for example, are obtained from the same strains. The reader will realize that applying the experimental results to other species, say humans, will compound the difficulties even further. In all of this subject matter knowledge is extremely important (see Chapters 1 and 2).

9.8 INCOMPLETE BLOCK DESIGNS

9.8.1 General Notion of Designs with Incomplete Blocks

In considering the RCBD and the GRBD we have assumed that the blocks contain enough homogeneous EUs so that each treatment can be applied once (for the RCBD) or $r(> 1)$ times (for the GRBD) in each block. Since the EUs in a block are rarely, if ever, homogeneous, large block sizes may be associated with large unit variances, σ_u^2 . As a consequence, the precision of the experiment, that is, its sensitivity to detect treatment differences may be adversely affected. It is an empirical fact that, generally, smaller blocks are less heterogeneous than larger blocks. Hence the possibility of using “small” blocks needs to be explored and should be given consideration in designing an experiment.

Often the experimenter is not given much of a choice in choosing an error-control design when blocks arise quite naturally with only few experimental units. As an extreme case, blocks of size two present themselves commonly, for example identical twins, half-leaves, the two sides of the body of an individual, such as both arms. Even litters of mice, parts of a field, or a batch of raw material may not have enough EUs to accommodate all treatments, especially if the number of treatments is large (examples of this we shall encounter when we discuss factorial experiments; see Chapter 11).

These situations give rise to what are called *incomplete blocks*, and the corresponding error-control designs are referred to as *incomplete block designs*. The question of how one should assign the treatments to the EUs in such blocks becomes then an important one. Obviously, different arrangements are possible, and some may be better than others. In this section we shall discuss, in general terms, several types of incomplete block designs. This is meant as an overview to give the reader some familiarity with the existence and nature of such designs. Most of the technical details will be deferred to Chapters II. 1–6.

The general situation is as follows: We have t treatments and b blocks; the i th block has k_i EUs ($i = 1, 2, \dots, b$) and the l th treatment is replicated r_l times ($l = 1, 2, \dots, t$). This implies obviously

$$\sum_{i=1}^b k_i = \sum_{l=1}^t r_l \quad (9.99)$$

and this number is denoted by n , the total number of EUs. From our earlier discussion it is also clear that not all treatments can occur equally often in each block, indeed in most situations not every treatment occurs in every block. The actual treatment-block

arrangement is characterized by the so-called *incidence matrix* $\mathbf{N} = (n_{li})$ which is a $t \times b$ matrix with elements

$$n_{li} = \text{number of times treatment } l \text{ occurs in block } i.$$

As a consequence we have

$$\sum_l n_{li} = k_i \quad (i = 1, 2, \dots, b)$$

and

$$\sum_i n_{li} = r_l \quad (l = 1, 2, \dots, t).$$

The arrangement of the treatments in blocks cannot be done haphazardly, since this may lead to undesirable properties of the design. An important property of a design is that of connectedness (see Section 4.12.4) which allows us to estimate all simple treatment comparisons of the form $\tau_l - \tau_{l'}$. The designs described in the following subsections possess this property and, in addition, have some other desirable features.

Before turning to these designs, however, we shall make a few general remarks about the analysis of incomplete block designs. Denoting the m -th observation for the l -th treatment in the i -th block by y_{ilm} , we write, assuming unit-treatment additivity in the broad sense,

$$y_{ilm} = \mu + \beta_i + \tau_l + e_{ilm} \quad (9.100a)$$

with $i = 1, 2, \dots, b$; $l = 1, 2, \dots, t$; $m = 1, 2, \dots, n_{il}$, or, in matrix notation,

$$\mathbf{y} = \mathbf{J}\mu + \mathbf{X}_\beta\boldsymbol{\beta} + \mathbf{X}_\tau\boldsymbol{\tau} + \mathbf{e} \quad (9.100b)$$

(see model (4.2) in Section 4.3.2). This model has been referred to earlier as a three-part linear model (see Section 4.9) or as a two-way classification model (see Sections 4.10 and 4.13.3).

The general form of the normal equations (NE) for estimating the effects in (9.100) is given in Section 4.13.3. To solve these equations it is convenient and informative to reduce the NE to a set of linear equations involving only the τ_l ($l = 1, 2, \dots, t$) by *absorbing* the equations for μ and the β_i into the equations for the τ_l . This leads to the so-called *reduced normal equations* (RNE) (for details see section II.1.3). In matrix notation the RNE can be written as

$$(\mathbf{R} - \mathbf{N}\mathbf{K}^{-1}\mathbf{N}')\hat{\boldsymbol{\tau}} = \mathbf{T} - \mathbf{N}\mathbf{K}^{-1}\mathbf{B}, \quad (9.101)$$

where

$$\mathbf{R} = \begin{pmatrix} r_1 & & & \\ & r_2 & & \\ & & \ddots & \\ 0 & & & r_t \end{pmatrix} \text{ and } \mathbf{K} = \begin{pmatrix} k_1 & & & \\ & k_2 & & \\ & & \ddots & \\ 0 & & & k_b \end{pmatrix}$$

Table 9.13 ANOVA for Incomplete Block Design

Source	d.f.	SS*	E(MS)
$\mathbf{X}_\beta \mathcal{J}$	$b - 1$	$\sum_{ij}^b \frac{B_i^2}{k_i} - \frac{G^2}{n}$	
$\mathbf{X}_\tau \mathcal{J}, \mathbf{X}_\beta$	$t - 1$	$\sum_{l=1}^t \hat{\tau}_l Q_l$	$\sigma_e^2 + \frac{\boldsymbol{\tau}' \mathbf{C} \boldsymbol{\tau}}{t - 1}$
Error	$n - b - t + 1$	Difference	σ_e^2
Total	$n - 1$	$\sum_{ijm} y_{ijm}^2 - \frac{G^2}{n}$	

*G=grand total, n=total number of observations

and $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_t)'$, $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_t)'$, $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_B)'$ with $\mathbf{T}_l = \sum_{i,m} y_{ilm} = l$ -th treatment total, $B_i = \sum_{l,m} y_{ilm} = i$ -th block total. In a more succinct form the RNE in (9.101) are generally written as

$$\mathbf{C} \hat{\boldsymbol{\tau}} = \mathbf{Q} \quad (9.102)$$

with \mathbf{C} representing the coefficient matrix for $\hat{\boldsymbol{\tau}}$ and \mathbf{Q} the right-hand side of the RNE. The so-called \mathbf{C} -matrix above is also referred to as the *information matrix* as it contains all the information concerning the properties of the underlying design.

The general form of the analysis of variance for an incomplete block design is given in Table 9.13. More precisely, the sums of squares given in Table 9.13 are the sequential sums of squares for the ordered model (9.100a,b) (Type I SS in SAS terminology). This is reflected in the expressions for the sources of variation, a notation established in Section 4.10.

In the terminology of block designs the sum of squares $SS(\mathbf{X}_\beta | \mathcal{J})$ is called the $SS(\text{Blocks ignoring treatments})$ and $SS(\mathbf{X}_\tau | \mathcal{J}, \mathbf{X}_\beta)$ is called $SS(\text{Treatments adjusted for blocks})$. For other details we refer to Section II.1.3.6.

We shall now turn our attention to some specific classes of incomplete block designs which can be described by the specific form of \mathbf{C} in (9.102).

9.8.2 Balanced Incomplete Block Designs

The *balanced incomplete block design* (BIBD) is an *equireplicate* (that is, all $r_l = r$), *proper* (that is, all $k_i = k$), *binary* design (that is, $n_{li} = 1$ or 0) introduced by Yates (1936). It has the additional and most important property that every pair of treatments

occurs together in a block the same number of times, this number being denoted by λ . We shall refer to such a design as BIBD $(t, b, k, r; \lambda)$ indicating thus the parameters of the design. For (9.99) we then have

$$tr = bk = n$$

and

$$\begin{aligned} \sum_i n_{li} &= k && \text{for every } i \\ \sum_i n_{li} &= r && \text{for every } l. \end{aligned}$$

We also have the following relationship between the parameters

$$\lambda(t-1) = r(k-1). \quad (9.103)$$

The validity of this relationship can be seen as follows: Consider one treatment, say l^* . Since l^* occurs exactly λ times together with the remaining $t-1$ treatments, the number of EUs occupied by those treatments in the blocks in which l^* occurs must be $q = \lambda(t-1)$. On the other hand, since l^* occurs in r blocks and each block has k EUs, the same number q is equal to $r(k-1)$.

The relationship (9.103) implies

$$\lambda = \frac{r(k-1)}{t-1}. \quad (9.104)$$

Since λ must be an integer it is clear that a BIBD does not exist for all values of t , k , and r . Even for values of t , k , and r yielding an integer λ , a BIBD may not exist. In fact there exists only a limited number of BIBDs in the useful parameter range. A (incomplete) list of actual plans is given by Cochran and Cox (1957). Raghavarao (1971) and Mathon and Rosa (1966) provide a more complete list of parameters of existing designs (see also Chapter II.3.4 for a listing of BIBDs for $t \leq 25$ and $k \leq 11$).

We give an example of a BIBD below.

EXAMPLE 9.4: We consider BIBD $(6, 10, 3, 5; 2)$ which can be written as follows, each triplet representing the treatments in a block (before randomization):

$$\begin{array}{ccc} 1 & 2 & 5 \\ 1 & 2 & 6 \\ 1 & 3 & 4 \\ 1 & 3 & 6 \\ 1 & 4 & 5 \end{array} \quad \begin{array}{ccc} 2 & 3 & 4 \\ 2 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 5 & 6 \\ 4 & 5 & 6 \end{array}$$

Equivalently, this design can be expressed in terms of the incidence matrix \mathbf{N} as

$$\mathbf{N} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

It can be seen easily from \mathbf{N} that pairs of 1s occur exactly twice, according to $\lambda = 2$, that is, $\sum_{i=1}^b n_{li} n_{l'i} = \lambda = 2$ for every pair $l, l' (l \neq l')$. The way we use this design is to randomly assign each triplet to a block and then randomly assign the treatments to the EUs within a block, using independent randomizations for the b blocks. \square

The analysis of observations from a BIBD is based on model (9.100a) which we now write simply as

$$y_{il} = \mu + \beta_i + \tau_l + e_{il} \quad (9.105)$$

assuming unit-treatment additivity in the broad sense using least squares analysis and the ANOVA in Table 9.13 with $k_i = k$ for all i . One important point here is that if we rewrite (9.105) in matrix form as

$$\mathbf{y} = \mathbf{J}\mu + \mathbf{X}_\beta\beta + \mathbf{X}_\tau\tau + \mathbf{e}$$

then we find for the BIBD (and the other incomplete block designs in this chapter) that

$$SS(\mathbf{X}_\tau | \mathbf{J} \mathbf{X}_\beta) \neq SS(\mathbf{X}_\tau | \mathbf{J})$$

(see Section 4.11). This means that for incomplete block designs model (9.105) is a *nonorthogonal model* or, expressed alternatively, incomplete block designs are *non-orthogonal designs*. As a further consequence of the incomplete block arrangement comparisons among treatments can no longer be accomplished by comparing treatment means. Rather, the comparisons are made using LS means (see Chapters II.1 and 2), because “adjustments” have to be made since not every treatment occurs in every block (this applies to all incomplete block designs).

For the BIBD the \mathbf{C} -matrix of (9.101) takes on the form

$$\mathbf{C} = \left(r \frac{k-1}{k} + \frac{\lambda}{k} \right) \mathbf{I} - \frac{\lambda}{k} \mathbf{J}\mathbf{J}'$$

(see Section II.2.4.1) with a generalized inverse (see Section 4.4.4) given by

$$\begin{aligned} \mathbf{C}^- &= \frac{k}{r(k-1) + \lambda} \mathbf{I} \\ &= \frac{1}{r - \frac{\lambda}{k}} \mathbf{I}. \end{aligned} \quad (9.106)$$

It then follows that for a simple treatment comparison $\hat{\tau}_l - \hat{\tau}_{l'}$ we obtain

$$\begin{aligned} \text{var}(\hat{\tau}_l - \hat{\tau}_{l'}) &= \frac{2}{r - \frac{\lambda}{k}} \sigma_e^2 \\ &= \frac{2}{rE} \sigma_e^2, \end{aligned} \quad (9.107)$$

where, using (9.104),

$$E = \frac{t(k-1)}{k(t-1)} \quad (9.108)$$

is referred to as the *efficiency factor* of the BIBD. It follows from (9.108) that $E < 1$. This would imply that a BIBD is less efficient than the corresponding RCBD with the same number of replications r . This ignores, however, the fact that usually

$$\sigma_{e(\text{BIBD})}^2 < \sigma_{e(\text{RCBD})}^2.$$

Hence a fair comparison of $\text{var}(\hat{\tau}_l - \hat{\tau}_{l'})_{\text{BIBD}}$ and $\text{var}(\hat{\tau}_l - \hat{\tau}_{l'})_{\text{RCBD}}$ would depend on the relationship between $\sigma_{e(\text{BIBD})}^2/E$ and $\sigma_{e(\text{RCBD})}^2$. Such information may not be available, in particular if a RCBD is not an available option. An estimate of σ_e^2 is obtained from the ANOVA table (Table 9.13) as $\hat{\sigma}_e^2 = \text{MS}(E) = \text{MS}(\mathbf{I}|\mathbf{J}\mathbf{X}_\beta\mathbf{X}_\tau)$. A test of the hypothesis $H_0: \tau_1 = \tau_2 = \dots = \tau_t = 0$ can also be derived from the ANOVA table, using $F = \text{MS}(\mathbf{X}_\tau|\mathbf{J}\mathbf{X}_\beta)/\text{MS}(\mathbf{I}|\mathbf{J}\mathbf{X}_\tau\mathbf{X}_\beta)$ with $t - 1$ and $n - b - t + 1$ d.f.

9.8.3 Balanced Treatment Incomplete Block Designs

In many kinds of experiments it is of primary importance to compare several treatments with an established procedure or a control, whereas the comparisons among the treatments are only of secondary importance or of no importance at all. As an example consider the efficacy of several drugs as compared to a placebo. The comparisons among the drugs may not be important as they control different side effects. Another example is mentioned by Pearce (1983): In a spraying experiment with insecticides it is useful to leave some EUs unsprayed to provide evidence that the infestation is present. Of primary interest then is to see to what extent the insecticides control the pest and comparisons among them may only be of secondary interest. Even though BIBDs can be used in these situations they are usually not the most appropriate designs since they consider all treatments as equally important. To emphasize and take into account the specific situations discussed above Pearce (1960) considered designs through supplementation and in a more systematic and comprehensive approach Bechhofer and Tamhane (1981, 1983) introduced and developed *balanced treatment incomplete block designs* (BTIBD).

Suppose we have one control treatment and t test treatments, denoted by 0, and $l = 1, 2, \dots, t$, respectively, with b blocks of size k ($k < t + 1$). An incomplete block design with these parameters is then called a BTIBD if for every l

$$\text{var}(\hat{\tau}_0 - \hat{\tau}_l) = \alpha^2 \sigma_e^2 \quad (9.109)$$

and for every pair l, l' ($l \neq l'$)

$$\text{cov}(\hat{\tau}_0 - \hat{\tau}_l, \hat{\tau}_0 - \hat{\tau}_{l'}) = \rho \alpha^2 \sigma_e^2, \quad (9.110)$$

where α and ρ depend on the design employed. Bechhofer and Tamhane (1981) show that a necessary and sufficient condition for an incomplete block design to be a BTIBD is that every test treatment occurs together λ_0 times with the control in the same block and any pair of test treatments occur together λ_1 times in the same block. In terms of the elements of the incidence matrix $\mathbf{N} = (n_{ji})$ ($j = 0, 1, 2, \dots, t; i = 1, 2, \dots, b$) these conditions can be written as

$$\sum_{i=1}^b n_{0i} n_{li} = \lambda_0 \quad (l = 1, 2, \dots, t)$$

and

$$\sum_{i=1}^b n_{li} n_{l'i} = \lambda_1 \quad (l, l' = 1, 2, \dots, t; l \neq l').$$

We denote such a design by BTIBD $(t, b, k; \lambda_0, \lambda_1)$. In order for all $\tau_0 - \tau_l (l = 1, 2, \dots, t)$ to be estimable we obviously require $\lambda_0 > 0$.

To illustrate this type of design we consider three examples.

EXAMPLE 9.5: BTIBD $(4, 6, 3; 3, 1)$ is given by

0	1	2
0	1	3
0	1	4
0	2	3
0	2	4
0	3	4

We note that this design is derived by supplementing each block in the BIBD $(4, 6, 2, 3; 1)$ with the control 0. As a consequence 0 is replicated $r_0 = 6$ times, whereas the test treatments are replicated $r = 3$ times. \square

EXAMPLE 9.6: BTIBD $(4, 7, 3; 2, 2)$ is given by

0	1	2
0	1	4
0	2	4
0	0	3
1	2	3
1	3	4
2	3	4

We note here that this design is not a binary design since $n_{04} = 2$. Also, $r_0 = 5, r = 4$. \square

EXAMPLE 9.7: Another BTIBD $(4, 7, 3; 2, 2)$ is given by

0	1	3
0	1	4
0	2	3
0	2	4
1	2	3
1	2	4
3	4	4

This is a design with unequal replication for the test treatments, that is, $r_0 = r_1 = r_2 = r_3 = 4, r_4 = 5$, where r_l denotes the number of replications for treatment l . \square

A list of BTIBDs is provided by Bechhofer and Tamhane (1985) together with their properties and methods of construction.

Least squares analysis based on model (9.100a) yields

$$\alpha^2 = \frac{k(\lambda_0 + \lambda_1)}{\lambda_0(\lambda_0 + t\lambda_1)}$$

for (9.109) and

$$\rho = \frac{\lambda_1}{\lambda_0 + \lambda_1}$$

for (9.110). For the designs in Examples 9.6 and 9.7 above it follows then that they have the same α^2 and ρ . These designs are said to be equivalent and considerations other than statistical may have to be used to decide which design to use in a practical situation.

For a more extensive discussion of various types of BTIBDs see Section II.6.5.

9.8.4 Partially Balanced Incomplete Block Designs

We have mentioned earlier that BIBDs exist only for a limited number of parameters and often with a large number of blocks. To provide practical alternatives Bose and Nair (1939) developed a large class of incomplete block designs, referred to as *partially balanced incomplete block designs* (PBIBD). Recall that the important property of BIBDs is that simple treatment comparisons are estimated with the same variance. An obvious relaxation of this requirement is to search for designs which allow two types of variances for all $t(t-1)/2$ simple treatment comparisons. This property is achieved by an important subclass of all PBIBDs, namely the so-called 2-associate class PBIBDs. A list of such designs is given by Clatworthy (1973) and more details about PBIBDs are provided in Chapters II. 4 and 5. We shall give here just one example of such a design to give the reader some insight into the nature of these designs.

EXAMPLE 9.8: Suppose we have $t = 6$ treatments and blocks of size $k = 4$. From the list of available BIBDs we see that $b = 15$ blocks are needed for a BIBD with these parameters. Suppose we do not have 15 blocks available and hence need an alternative design. From Clatworthy (1973) we obtain the following design (his design $S1$) with only $b = 3$ blocks and $r = 2$ replications for each treatment (each row represents a block):

1	4	2	5
2	5	3	6
3	6	1	4

We notice, by inspection, that several pairs of treatments occur together twice in the same block (1 and 4, 2 and 5, 3 and 6) and the remaining pairs occur together once in the same block. This leads, for each treatment, to a classification of the remaining treatments into what are called *associate classes*. In this case this is done as follows: Write the treatments in a rectangular array

1	4
2	5
3	6

and declare any two treatments occurring in the same row to be 1. associates and any two treatments not occurring in the same row to be 2. associates. This association scheme then leads to the following classification:

Treatment	1. Associates	2. Associates
1	4	2, 3, 5, 6
2	5	1, 3, 4, 6
3	6	1, 2, 4, 5
4	1	2, 3, 5, 6
5	2	1, 3, 4, 6
6	3	1, 2, 4, 5

Any two treatments which are 1. associates, for instance, 1 and 4, occur together λ_1 times in the same block, and any two treatments which are 2. associates, for instance, 1 and 2, occur together λ_2 times in the same block. We saw already that for our design given above we have $\lambda_1 = 2, \lambda_2 = 1$. If in general we write a PBIBD as PBIBD $(t, b, k, r; \lambda_1, \lambda_2)$ we have in particular here a PBIBD $(6, 3, 4, 2; 2, 1)$.

Because of the fact that $\lambda_1 \neq \lambda_2$ (if $\lambda_1 = \lambda_2$ we would have a BIBD), we have that any two treatments which are 1. associates are compared with one variance, v_1 say, and any two treatments which are 2. associates are compared with another variance, v_2 say, for example,

$$\text{var}(\hat{\tau}_1 - \hat{\tau}_4) = v_1$$

and

$$\text{var}(\hat{\tau}_1 - \hat{\tau}_2) = v_2.$$

The average variance for treatment comparisons then is

$$\text{av. var} = \frac{n_1 v_1 + n_2 v_2}{t - 1},$$

where n_i is the number of i th associates ($i = 1, 2$) with $n_1 + n_2 = t - 1$. Analogously we can define two efficiency factors E_1 and E_2 by

$$v_i = \frac{2}{r E_i} \sigma_e^2 \quad (i = 1, 2)$$

and the overall efficiency factor E as

$$E = \frac{n_1 E_1 + n_2 E_2}{t - 1}.$$

These efficiency factors are useful for comparisons of competing designs. They are given together with other relevant parameters in Clatworthy's (1973) tables. For our example we find (see Chapter II.4) $v_1 = \sigma_e^2, v_2 = 1.16\sigma_e^2$, $\text{av. var} = 1.14\sigma_e^2, E_1 = 1.00, E_2 = .86, E = .88$. \square

The list of available and practically useful PBIBDs is rather extensive. Since the construction of PBIBDs requires in general a certain amount of mathematical machinery (see Chapter II.5) this list is also very convenient to have as a tool for finding suitable designs for a given experimental situation. As we shall discuss in later chapters, the notion of PBIBDs in their general form is quite fundamental in other aspects of experimental design as well. This includes PBIBDs with more than two associate classes (see also Section II.4.6).

9.8.5 Extended Block Designs

In the preceding sections we have considered situations in which the block size k is less than the number of treatments t . It is, of course, also possible to encounter experimental situations in which $\gamma t < k < (\gamma + 1)t$, with γ being a positive integer. Since the case $\gamma = 1$ is of particular importance and interest we shall confine ourselves to this case here, but extensions should be quite obvious.

Suppose then we have t treatments, b blocks of size k with $t < k < 2t$. It is obvious then that each treatment can occur in each block once but not twice. A reasonable approach would be to assign in each block t EUs to the t treatments and then fill in each block the remaining $k - t$ EUs in an appropriate manner in concordance with the objective of the experiment. It is difficult to give any general rules, so we shall give only a few examples to illustrate the general idea:

EXAMPLE 9.9: Suppose $k = t + 1$. If one treatment is of special importance, for example, a standard with which all other treatments should be compared, then it might be appropriate to assign that treatment to the additional EU in each block. Such a design would then have the same properties as a BTIBD (see Section 9.8.3). If no treatment is of special importance, then having equal or nearly equal replication for each treatment would seem to be appropriate. □

EXAMPLE 9.10: Suppose $t + 1 < k < 2t$. One possible approach is to adjoin to the RCBD part of the overall design one of the incomplete block designs discussed in the previous sections. Consider the case $t = 4, k = 6, b = 6$. Here we could combine a RCBD with a BIBD (4, 6, 2, 3; 1) so that the final design looks like this (before randomization)

Block	Treatments	
1	1 2 3 4	1 2
2	1 2 3 4	1 3
3	1 2 3 4	1 4
4	1 2 3 4	2 3
5	1 2 3 4	2 4
6	<u>1 2 3 4</u>	<u>3 4</u>
	RCBD	BIBD

□

EXAMPLE 9.11: Suppose $t = 4, k = 7, b = 6$ and treatment 1 is of special importance. One possible design then is the following (before randomization)

Block	Treatments						
1	1	2	3	4	1	2	1
2	1	2	3	4	1	3	1
3	1	2	3	4	1	4	1
4	1	2	3	4	2	3	1
5	1	2	3	4	2	4	1
6	1	2	3	4	3	4	1
RCBD				BIBD			
				BTIBD			

□

An important aspect of *extended block designs* is that they allow easy separation of block-treatment interaction and error if that is desirable. For Example 9.11 above the partitioning of the total number of d.f. in the ANOVA is as follows:

Source	d.f.
Blocks	5
Treatments	3
$B \times T$	15
Error	18
Total	41

The error d.f. arise, of course, from comparisons among EUs treated alike in a block.

9.8.6 Some General Remarks

The classes of designs mentioned in the previous sections represent only a fraction of existing incomplete block designs. We have included them here because the designs in these classes (i) have certain structures which can be explained quite easily, (ii) are, for the most part, of practical value, (iii) are easily analyzed, and (iv) serve very often as foundation or building blocks for other designs.

The fact that these designs have a structure, that is, certain combinatorial and statistical properties, leads to a fairly easy, albeit nonorthogonal analysis. This was certainly the major reason for the development of these designs. With today’s computing facilities this is no longer of great importance, but useful nonetheless. Other designs have been developed which do not have the kind of structure as BIBDs or PBIBDs but which still have certain properties. Among those are the pairwise balanced and variance balanced designs (see John, 1964; Hedayat and Federer, 1974). These are for the most part designs with unequal block sizes, unequal numbers of replications and unequal

concurrences of treatments in the same block. Even though these designs possess certain combinatorial properties they are much more difficult to describe in general and even more difficult to list.

With respect to designs with unequal block sizes, care must be exercised in their use. One of the basic assumptions for the analysis of block designs is that of equal variances within blocks. As we mentioned earlier, the variance tends to increase as the block size increases. As long as the block sizes are not too unequal, this should not be a major problem. Situations of this type include different litter sizes, different numbers of leaves per plant, different block sizes due to irregular shape of experimental field, and so forth.

The structure of the designs discussed ensures also that all treatment contrasts can be estimated, in particular all treatment differences can be estimated. We refer to such designs as *connected designs*, whereas designs which do not have this property are called *disconnected designs* (see also Section 4.12.4). To illustrate the generally undesirable property of disconnectedness, consider the following example.

EXAMPLE 9.12: Consider the incomplete block design with $t = 5$ and $b = 4$ as given by its incidence matrix

$$\mathbf{N} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We see immediately that the treatments fall into two sets such that treatments in the first set, $\{1, 2, 3\}$, do not occur together in the same block with treatments of the second set, $\{4, 5\}$. As a consequence, functions of the form $\tau_{l'} - \tau_{l''}$ with $l' = 1, 2, 3$ and $l'' = 4, 5$ cannot be estimated unbiasedly since, using model (9.105) we cannot “eliminate” the block effects. Expressed differently, there does not exist a linear combination of the observations, $\sum_{i,l} a_{il} y_{il}$ say, such that its expected value is $\tau_{l'} - \tau_{l''}$ for some l', l'' . Had the design instead been of the form

$$\mathbf{N} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

with treatment 3 occurring also in block 4 we would have a connected design. Now all treatment differences can be estimated as can be seen simply by looking at the individual blocks and the treatment comparisons estimable within them and using the fact that if $\tau_l - \tau_{l'}$ and $\tau_{l'} - \tau_{l''}$ are estimable then also

$$(\tau_l - \tau_{l'}) + (\tau_{l'} - \tau_{l''}) = \tau_l - \tau_{l''}$$

is estimable. This brief discussion (for more detail see Chapter II.1) is meant to point out that the assignment of treatments to blocks should generally not be done haphazardly but always with a view towards achieving connectedness. \square

Generally, block designs are of major importance in experimental work. Careful attention must be given to the availability or formation of blocks and an appropriate arrangement of the treatments must be selected according to the objectives of the experiment. Often one blocking factor is not sufficient (we have given examples in Section 9.6.7) and in other situations “blocking in different directions” may be called for. Examples of that will be discussed in the following chapter.

9.9 SYSTEMATIC BLOCK DESIGNS

In our discussion of the various forms of block designs we have emphasized the random assignment of treatments to EUs, using independent randomizations for each block. The reason for such random assignment is, of course, to avoid any bias in the treatment comparisons (see Chapter 2). There exist, however, situations where it may be advantageous to employ allocations of the treatments to EU's other than random allocations. One such situation presents itself if the unit contributions U_{ij} [see (9.1)] exhibit some sort of smooth trend within the i th block or if the status of the EUs in a block change in a gradient fashion as the treatments are applied sequentially. The question then arises: Can we utilize this knowledge and allocate the treatments such that this leads to “improved” estimation of treatment comparisons vis-a-vis the situation where this information is being ignored?

9.9.1 Dealing with Trends

Considering a complete block design, we assume that we can express the trend in terms of a polynomial of known degree, say p , with $p < t$, of some characteristic of the EUs, denoted by x . One obvious way to proceed is to use random allocation as in the usual RCBD and use the information about the trend as supplementary information in conjunction with an analysis of covariance model. Let y_{ij} denote the observation for the j th EU in the i th block. Assume that the trend is the same in each block, that is, the trend is a function only of x_j ($j = 1, 2, \dots, b$). We can then write

$$y_{ij} = \mu + \beta_i + \sum_{k=1}^t \delta_{ij}^k \tau_k + \sum_{l=1}^p \gamma_l^* x_j^l + e_{ij}^*$$

or, more conveniently, in terms of orthogonal polynomials (see Section 7.4)

$$y_{ij} = \mu + \beta_i + \sum_k \delta_{ij}^k \tau_k + \sum_l \gamma_l P_l(x_j) + e_{ij}^* \quad (9.111)$$

where $\delta_{ij}^k = 1$ if treatment k is applied to the j th EU in the i th block, and 0 otherwise. This model can be simplified even further by taking $x_j = j$ and hence writing (9.111) as

$$y_{ij} = \mu + \beta_i + \sum_k \delta_{ij}^k \tau_k + \sum_l \gamma_l P_l(j) + e_{ij}^* \quad (9.112)$$

We rewrite (9.112) in matrix notation as

$$\mathbf{y} = \mathbf{J}\mu + \mathbf{X}_\beta\beta + \mathbf{X}_\tau\tau + \mathbf{X}_\gamma\gamma + \mathbf{e}^*. \quad (9.113)$$

We can analyze the data, using model (9.113), in the usual fashion (see Section 9.4 and Chapter 4). The important point to keep in mind here is that the sum of squares for treatments is of the form $SS(\mathbf{X}_\tau | \mathbf{J} \mathbf{X}_\gamma)$ as compared to $SS(\mathbf{X}_\tau | \mathbf{J})$ for the RCBD without covariate, and that for model (9.113) generally

$$SS(\mathbf{X}_\tau | \mathbf{J} \mathbf{X}_\gamma) \neq SS(\mathbf{X}_\tau | \mathbf{J}).$$

Cox (1951) has pointed out that if a model of the form (9.113) holds the method just described leads to a loss of information with regard to treatment comparisons.

9.9.2 Trend-free Designs

An alternative method then is to use some systematic arrangement of the treatments in a block. Such possibilities were considered first, although in a somewhat different context, by e.g., Neyman (1929) and Cox (1951, 1952). One possibility, for example is, to repeat the treatments in the same order (if the EUs are layed out along a line), for example, $T_1, T_2, T_3, T_1, T_2, T_3, \dots$; or in a mirror image fashion, for example, $T_2, T_2, T_1, T_1, T_1, T_1, T_2, T_2$ for a linear trend. The idea is to construct what are now referred to as *trend-free designs*, where a design is considered to be trend-free if, generally speaking, the sum of squares due to treatments is not affected by the covariate, that is, by the model for the trend.

For complete blocks the design is trend-free if, in our earlier notation,

$$SS(\mathbf{X}_\tau | \mathbf{J} \mathbf{X}_\gamma) = SS(\mathbf{X}_\tau | \mathbf{J}).$$

More generally, for incomplete blocks of size $k(\leq t)$ and $p < k$, Bradley and Yeh (1980) give the following definition: A block design modeled by (9.113) is trend-free relative to the trend in the model if

$$SS(\mathbf{X}_\tau | \mathbf{J} \mathbf{X}_\beta \mathbf{X}_\gamma) = SS(\mathbf{X}_\tau | \mathbf{J} \mathbf{X}_\beta). \quad (9.114)$$

They show that a necessary and sufficient condition for a block design to be trend-free is that

$$\mathbf{X}'_\tau \mathbf{X}_\gamma = \mathbf{O}. \quad (9.115)$$

The construction of such designs is not straightforward. For the important case of a linear trend the existence of trend-free block designs has been shown. For the complete block design Yeh and Bradley (1983) prove that a necessary and sufficient condition is that either b is even, or both b and t are odd, with $b \geq 3$.

EXAMPLE 9.13: For $t = 7$ and $b = 3$ the following design is trend-free:

	Treatments						
Block 1	1	2	3	4	5	6	7
Block 2	6	4	2	7	5	3	1
Block 3	7	5	3	1	6	4	2
$P_1(j)$	-3	-2	-1	0	1	2	3

This can be verified easily by adding for each treatment the coefficients of the orthogonal polynomial of degree 1 (given below the design) corresponding to the positions in which the treatment occurs. They add to zero, hence satisfying (9.115). \square

For the equireplicate, proper, binary block design Stufken (1988) has shown that a linear trend-free block design exists if and only if $r(k + 1)$ is even with $r \geq 2$.

EXAMPLE 9.14: The following BIBD $(5, 10, 3, 6; 3)$ is trend-free (Bradley and Yeh, 1980):

	Treatments		
Block 1	1	2	3
Block 2	1	2	4
Block 3	1	2	5
Block 4	3	4	1
Block 5	3	5	1
Block 6	4	5	1
Block 7	4	2	3
Block 8	5	2	3
Block 9	5	2	4
Block 10	3	4	5
$P_1(j)$	-1	0	1

The reader can verify easily that (9.115) is satisfied. \square

It is quite obvious from this limited discussion that trend-free block designs do not exist for all situations and even if they do exist their construction is not obvious. For some construction methods and a generalization of the concept of trend-free to nearly trend-free designs we refer the reader to Yeh, Bradley and Notz (1985), and Bradley and Odeh (1988).

The assumption that the trend is the same in each block may not always be realistic. Allowing for different linear trends in the blocks Jacroux et al. (1995) and Jacroux (1998) have provided methods of constructing efficient and optimal (see Section II.1.13) designs. It is not surprising that many of these designs have as the basis a BIBD or PBIBD, as shown in the following example.

EXAMPLE 9.15: (Jacroux, 1998): For a trend-free design with $t = 6$, $b = 9$, $k = 3$ the following PBIBD with $t = 6$, $b = 9$, $k = 2$:

	Block								
	1	2	3	4	5	6	7	8	9
Treatments	1	1	1	2	2	2	3	3	3
	4	5	6	4	5	6	4	5	6

is augmented by replicating the first row to form the final design for $t = 6$, $b = 9$, $k = 3$:

	Block								
	1	2	3	4	5	6	7	8	9
	1	1	1	2	2	2	3	3	3
	4	5	6	4	5	6	4	5	6
Treatments	1	1	1	2	2	2	3	3	3

It is easy to see how this design is trend-free with respect to possibly different linear trends within blocks. Obviously, it is also trend-free for a common trend, as we can verify easily following arguments given earlier. \square

9.10 EXAMPLES USING SAS®

In this section we illustrate some of the analysis procedures described in this chapter, with numerical examples. In each case we describe an experimental situation and give a data set. The analysis is carried out with the help of SAS procedures (SAS Institute, Inc. 2002-2003), in particular SAS PROC GLM and SAS PROC MIXED. We shall make some comments about the input statements and the output as a link to the developments in this chapter.

EXAMPLE 9.16: Consider an experiment, using an RCBD, to study weight gain in rabbits due to five different diets: 1 = standard, 2 = 10 protein added, 3 = 20% protein added, 4 = additive *A*, 5 = additive *B*. We use six litters of rabbits, each litter containing five animals. The litters represent the blocks and the individual animals represent the EUs. The data are given in Table 9.14a.

We use SAS PROC GLM to analyze the data. The input statements are given in Table 9.14a: In addition to the ANOVA we perform Tukey's multiple comparison test with $\alpha = .10$ and we specify a complete set of orthogonal contrasts reflecting the structure of the treatments.

The results of the analysis are given in Table 9.14b, and we comment briefly on some aspects of the output:

- (i) Since the data set is balanced the Type I and III SS are identical and equal to those obtainable from Table 9.2.
- (ii) Since PROC GLM is a general linear models program that cannot distinguish between data from observational or intervention studies it automatically performs tests of significance for all effects specified in the model statement. We should, therefore, ignore the P -value for testing block (litter) effects. The P -value for diets (0.0282) indicates that there exist differences among the diets.
- (iii) Tukey's test at $\alpha = .10$ indicates a significant difference only between diet 1 and diet 5.

Table 9.14 Randomized Complete Block Design ($t = 5, b = 6$)

a) Input statements:

```
data weight;
input diet litter gain @@;
datalines;
1 1 57.0 1 2 55.0 1 3 62.1 1 4 74.5 1 5 86.7 1 6 42.0
2 1 64.8 2 2 66.6 2 3 69.5 2 4 61.1 2 5 91.8 2 6 51.8
3 1 70.7 3 2 59.4 3 3 64.5 3 4 74.0 3 5 78.5 3 6 55.8
4 1 68.3 4 2 67.1 4 3 69.1 4 4 72.7 4 5 90.6 4 6 44.3
5 1 76.0 5 2 74.5 5 3 76.5 5 4 86.6 5 5 94.7 5 6 43.2
;
run;

proc glm data=weight;
class diet litter;
model gain=litter diet;
means diet/Tukey alpha=.10;
lsmeans diet/stderr;
contrast '1 vs rest' diet 4 -1 -1 -1 -1;
estimate '1 vs rest' diet 4 -1 -1 -1 -1;
contrast '1 vs rest' diet 4 -1 -1 -1 -1;
contrast '2+3 vs 4+5' diet 0 1 1 -1 -1;
contrast '2 vs 3' diet 0 1 -1 0 0;
contrast '4 vs 5' diet 0 0 0 1 -1;
title1 'RANDOMIZED COMPLETE BLOCK DESIGN (t=5, b=6)';
title2 'ANALYSIS OF VARIANCE W/POST-HOC COMPARISONS';
run;

proc mixed data=weight;
class diet litter;
model gain=diet;
random litter;
lsmeans diet;
contrast '1 vs rest' diet 4 -1 -1 -1 -1;
title2 'ASSUMING RANDOM BLOCK EFFECTS';
run;
```

b) Output:

RANDOMIZED COMPLETE BLOCK DESIGN (t=5, b=6)			
ANALYSIS OF VARIANCE W/POST-HOC COMPARISONS			
The GLM Procedure			
Class Level Information			
Class	Levels	Values	
diet	5	1 2 3 4 5	
litter	6	1 2 3 4 5 6	

Table 9.14 (Continued)

Number of Observations Read 30					
The GLM Procedure					
Dependent Variable: gain					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	4915.622667	546.180296	15.54	<.0001
Error	20	702.752000	35.137600		
Corrected Total	29	5618.374667			
	R-Square	Coeff Var	Root MSE	gain Mean	
	0.874919	8.677219	5.927698	68.31333	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
litter	5	4438.014667	887.602933	25.26	<.0001
diet	4	477.608000	119.402000	3.40	0.0282
Source	DF	Type III SS	Mean Square	F Value	Pr > F
litter	5	4438.014667	887.602933	25.26	<.0001
diet	4	477.608000	119.402000	3.40	0.0282
Tukey's Studentized Range (HSD) Test for gain					
Alpha				0.1	
Error Degrees of Freedom				20	
Error Mean Square				35.1376	
Critical Value of Studentized Range				3.73641	
Minimum Significant Difference				9.042	
Means with the same letter are not significantly different.					
Tukey Grouping		Mean	N	diet	
A		75.250	6	5	
A					
B	A	68.683	6	4	
B	A				
B	A	67.600	6	2	
B	A				
B	A	67.150	6	3	
B					
B		62.883	6	1	

Table 9.14 (Continued)

Least Squares Means				
	diet	gain LSMEAN	Standard Error	Pr > t
	1	62.8833333	2.4199725	<.0001
	2	67.6000000	2.4199725	<.0001
	3	67.1500000	2.4199725	<.0001
	4	68.6833333	2.4199725	<.0001
	5	75.2500000	2.4199725	<.0001

Dependent Variable: gain

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
1 vs rest	1	221.1367500	221.1367500	6.29	0.0208
2+3 vs 4+5	1	126.5004167	126.5004167	3.60	0.0723
2 vs 3	1	0.6075000	0.6075000	0.02	0.8967
4 vs 5	1	129.3633333	129.3633333	3.68	0.0694

Parameter	Estimate	Standard Error	t Value	Pr > t
1 vs rest	-27.1500000	10.8224458	-2.51	0.0208

ASSUMING RANDOM BLOCK EFFECTS

The Mixed Procedure

Model Information

Data Set	WORK.WEIGHT
Dependent Variable	gain
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
diet	5	1 2 3 4 5
litter	6	1 2 3 4 5 6

Dimensions

Covariance Parameters	2
Columns in X	6
Columns in Z	6

Table 9.14 (Continued)

Iteration History						
Iteration	Evaluations	-2 Res Log Like	Criterion			
0	1	213.05776599				
1	1	185.03378213	0.00000000			
Convergence criteria met						
Covariance Parameter Estimates						
Cov Parm		Estimate				
litter		170.49				
Residual		35.1376				
Fit Statistics						
-2 Res Log Likelihood		185.0				
AIC (smaller is better)		189.0				
AICC (smaller is better)		189.6				
BIC (smaller is better)		188.6				
Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	F Value	Pr > F		
diet	4	20	3.40	0.0282		
Contrasts						
Label	Num DF	Den DF	F Value	Pr > F		
1 vs rest	1	20	6.29	0.0208		
Least Squares Means						
Effect	diet	Estimate	Standard Error	DF	t Value	Pr > t
diet	1	62.8833	5.8542	20	10.74	<.0001
diet	2	67.6000	5.8542	20	11.55	<.0001
diet	3	67.1500	5.8542	20	11.47	<.0001
diet	4	68.6833	5.8542	20	11.73	<.0001
diet	5	75.2500	5.8542	20	12.85	<.0001

- (iv) The reader can verify easily that the sum of the contrast SSs is equal to the diet SS.
- (v) The tests concerning the set of orthogonal contrasts indicates that the difference among the diets is due mainly to the difference between diet 1 and the average of the new diets ($P = 0.0208$), the estimated difference between the weight gains being 27.15/4 grams. \square

EXAMPLE 9.17: We consider the same experimental situation and data as given in Example 9.16, except that we now consider the litter (block) effects to be random effects (see Section 9.7.6). We use SAS PROC MIXED to analyze the data. The input statements are given in Table 9.14a and the output in Table 9.14b.

We provide the following comments:

- (i) The variance component estimates are obtained as $\hat{\sigma}_\beta^2 = 170.49$ and $\hat{\sigma}_e^2 = 35.14$, the latter being the same as for the fixed effects model in Example 9.16.
- (ii) Tests of hypotheses for diets and contrasts among diets are the same as in Example 9.16.
- (iii) The LS means are the same as those in Example 9.16, but the standard errors are larger, 5.82 vs. 2.42 in Example 9.16, reflecting the wider inference space. \square

EXAMPLE 9.18: This example describes an experiment using an RCBD with sub-sampling. Suppose we want to compare the effect of three exercise regimens, say no exercise and two different forms of exercise. We have five patients (subjects) and each subject performs all three exercises in random order (after appropriate resting periods). Immediately after the exercise the blood pressure is taken twice (one measurement right after the other). Suppose the data for the diastolic pressure are as given in Table 9.15a.

We use both PROC GLM and PROC MIXED. The main purpose of using PROC GLM is to obtain the ANOVA table as outlined in Table 9.3.

The input statements are provided in Table 9.15a and the output in Table 9.15b. We make the following comments:

- (i) For both PROC GLM and PROC MIXED we have to describe the experimental error in technical terms, which is formally the (negligible) subject-exercise interaction. In addition, for PROC GLM this term has to be identified explicitly for any tests concerning the exercise effects, that is, overall test in the ANOVA, multiple comparison tests, contrast tests, as well as for obtaining the standard error for LS means. In PROC MIXED this will be achieved automatically by declaring the subject-exercise interaction as a random effect.

Table 9.15 Randomized Complete Block Design with Subsampling

a) Input statements:

```
data pressure;
input subject exercise diast @@;
datalines;
1 1 126 1 1 129 1 2 137 1 2 135 1 3 135 1 3 136
2 1 134 2 1 138 2 2 140 2 2 145 2 3 141 2 3 139
3 1 120 3 1 119 3 2 130 3 2 134 3 3 130 3 3 129
4 1 137 4 1 134 4 2 147 4 2 144 4 3 143 4 3 147
5 1 123 5 1 123 5 2 136 5 2 135 5 3 134 5 3 136
;
run;

proc glm data=pressure;
class subject exercise;
model diast = subject exercise subject*exercise;
test h=exercise e=subject*exercise;
lsmeans exercise/stderr pdiff adjust=Tukey e=subject*exercise;
contrast '1 vs 2+3';
exercise 2 -1 -1/e=subject*exercise;
title1 'RANDOMIZED COMPLETE BLOCK DESIGN';
title2 'WITH SUBSAMPLING';
title3 '(t=3, b=5, n=2)';
run;

proc mixed data=pressure;
class subject exercise;
model diast subject exercise;
random subject*exercise;
lsmeans exercise/pdiff adjust=Tukey;
contrast '1 vs 2+3' exercise 2 -1 -1;
estimate '1 vs 2+3' exercise 1 -.5 -.5;
run;
```

b) Output:

RANDOMIZED COMPLETE BLOCK DESIGN	
WITH SUBSAMPLING	
(t=3, b=5; n=2)	
The GLM Procedure	
Class Level Information	
Class	Levels Values
subject	5 1 2 3 4 5
exercise	3 1 2 3
Number of Observations Read	
Number of Observations Used	
30	
30	

Table 9.15 (Continued)

The GLM Procedure					
Dependent Variable: diast					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	1541.466667	110.104762	28.48	<.0001
Error	15	58.000000	3.866667		
Corrected Total	29	1599.466667			
	R-Square	Coeff Var	Root MSE	diast Mean	
	0.963738	1.461633	1.966384	134.5333	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
subject	4	905.1333333	226.2833333	58.52	<.0001
exercise	2	591.2666667	295.6333333	76.46	<.0001
subject*exercise	8	45.0666667	5.6333333	1.46	0.2522
Source	DF	Type III SS	Mean Square	F Value	Pr > F
subject	4	905.1333333	226.2833333	58.52	<.0001
exercise	2	591.2666667	295.6333333	76.46	<.0001
subject*exercise	8	45.0666667	5.6333333	1.46	0.2522
Tests of Hypotheses Using the Type III MS for subject*exercise as an Error Term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
exercise	2	591.2666667	295.6333333	52.48	<.0001
Least Squares Means					
Adjustment for Multiple Comparisons: Tukey					
Standard Errors and Probabilities Calculated Using the Type III MS for subject*exercise as an					
Error Term					
exercise	diast LSMEAN	Standard Error	Pr > t	LSMEAN Number	
1	128.300000	0.750555	<.0001	1	
2	138.300000	0.750555	<.0001	2	
3	137.000000	0.750555	<.0001	3	

Table 9.15 (Continued)

Least Squares Means for effect exercise				
Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: diast				
i/j	1	2	3	
1		<.0001	<.0001	
2	<.0001		0.4726	
3	<.0001	0.4726		
Dependent Variable: diast				
Tests of Hypotheses Using the Type III MS for subject*exercise as an Error Term				
Contrast	DF	Contrast SS	Mean Square	F Value Pr > F
1 vs 2+3	1	582.8166667	582.8166667	103.46 <.0001
The Mixed Procedure				
Model Information				
Data Set	WORK.PRESSURE			
Dependent Variable	diast			
Covariance Structure	Variance Components			
Estimation Method	REML			
Residual Variance Method	Profile			
Fixed Effects SE Method	Model-Based			
Degrees of Freedom Method	Containment			
Class Level Information				
Class	Levels	Values		
subject	5	1 2 3 4 5		
exercise	3	1 2 3		
Iteration History				
Iteration	Evaluations	-2 Res Log Like		Criterion
0	1	112.23380945		
1	1	111.85203058		0.00000000
Convergence criteria met.				

Table 9.15 (Continued)

Covariance Parameter Estimates									
Cov Parm			Estimate						
subject*exercise			0.8833						
Residual			3.8667						
Type 3 Tests of Fixed Effects									
Effect	Num DF	Den DF	F Value	Pr > F					
subject	4	8	40.17	<.0001					
exercise	2	8	52.48	<.0001					
Estimates									
Label	Estimate	Standard Error	DF	t Value	Pr > t				
1 vs 2+3	-9.3500	0.9192	8	-10.17	<.0001				
Contrasts									
Label	Num DF	Den DF	F Value	Pr > F					
1 vs 2+3	1	8	103.46	<.0001					
Least Squares Means									
Effect	exercise	Estimate	Standard Error	DF	t Value	Pr > t			
exercise	1	128.30	0.7506	8	170.94	<.0001			
exercise	2	138.30	0.7506	8	184.26	<.0001			
exercise	3	137.00	0.7506	8	182.53	<.0001			
Differences of Least Squares Means									
Standard Differences of Least Squares Means									
Effect	exercise	_exercise	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
exercise	1	2	-10.0000	1.0614	8	-9.42	<.0001	Tukey	<.0001
exercise	1	3	-8.7000	1.0614	8	-8.20	<.0001	Tukey	<.0001
exercise	2	3	1.3000	1.0614	8	1.22	0.2555	Tukey	0.472

- (ii) The output shows that there exist significant differences among the types of exercise ($P < .00010$), but that exercises 2 and 3 are not significantly different from each other ($P = .47$).
- (iii) The standard errors for the LS means are the same for both PROC GLM and PROC MIXED, as they should be for balanced data (they may be different for unbalanced data due to difference estimation procedures for estimating variance components; see Section II.1.11.2).
- (iv) The estimates for σ_ϵ^2 and σ_η^2 are given by PROC MIXED as $\hat{\sigma}_\epsilon^2 = 0.8833$ and $\hat{\sigma}_\eta^2 = 3.8667$. \square

EXAMPLE 9.19: The setting for this experiment using an RCBD with a nested blocking structure and subsampling is described in Example 9.1 (we have changed the numbers of half-sib families (HSF) and blocks per location in order to save space). Suppose we have obtained the data (height in cm) given in Table 9.16.

In order to obtain the ANOVA table given in Table 9.6.8 we use SAS PROC GLM. The input statements and the results are given in Table 9.16. We also illustrate how to use SAS PROC MIXED with input statements given in Table 9.16a and the output in Table 9.16b. We make the following comments:

- (i) For PROC GLM we have to provide a technical expression for the experimental error, which in this case is equal to the HSF \times block (location) interaction. This term, that is, the corresponding MS is used to test hypotheses about HSF and HSF \times location interaction.
- (ii) In PROC MIXED HSF \times block(loc) is considered to be a random effect, and correct tests about HSF and HSF \times location interaction are performed automatically. Both tests are significant ($P < 0.0001$). A look at the loc \times HFS LS means shows that the interaction is codirectional. Hence the test about HSF (averaged over locations) seems appropriate.
- (iii) The SLICE option in the LS means input statement is one way to investigate a significant interaction, in particular, if the interaction turns out to be antidirectional. We have included it here to indicate this option to perform the ANOVA separately for each location and provide the F-test for HSF for each location. Note that Denominator DF=12 indicates that the pooled experimental error has been used. \square

EXAMPLE 9.20: We consider here a fertilizer study involving two small grain varieties. The fertilizer is nitrogen (N) at five increasing levels (by the same amount). The field experiment is laid out as a GRBD, with the varieties representing the blocks, and each nitrogen level being applied to two EUs for each variety, that is, we have $t = 5$, $b = 2$, $r = 2$. Suppose we obtain the yield data given in Table 9.17a.

We use SAS PROC GLM to analyze the data. The input statements and the output are given in Table 9.17. We make the following comments:

Table 9.16 RCBD with Nested Blocking Structure and Subsampling

a) Input statements:

```
data pine;
input loc block HSF height @@;
datalines;
1 1 1 210 1 1 1 221 1 1 2 252 1 1 2 260 1 1 3 197 1 1 3 190
1 2 1 222 1 2 1 214 1 2 2 265 1 2 2 271 1 2 3 201 1 2 3 210
1 3 1 220 1 3 1 225 1 3 2 271 1 3 2 277 1 3 3 205 1 3 3 204
1 4 1 224 1 4 1 231 1 4 2 270 1 4 2 283 1 4 3 211 1 4 3 216
2 1 1 178 2 1 1 175 2 1 2 191 2 1 2 193 2 1 3 182 2 1 3 179
2 2 1 180 2 2 1 184 2 2 2 198 2 2 2 201 2 2 3 183 2 2 3 190
2 3 1 189 2 3 1 183 2 3 2 200 2 3 2 195 2 3 3 197 2 3 3 205
2 4 1 184 2 4 1 192 2 4 2 197 2 4 2 204 2 4 3 192 2 4 3 190
;
run;

proc glm data=pine;
class loc block HSF;
model height=loc block(loc) HSF loc*block(loc);
test h=exercise e=subject*exercise;
title1 'RCBD WITH NESTED BLOCKING STRUCTURE';
title2 'AND SUBSAMPLING';
title3 '[t=3, b=8 (A=2, C=4), n=2]';
run;

proc mixed data=pine;
class loc block HSF;
model height=loc block(loc) HSF loc*HSF;
random HSF*block(loc);
lsmeans HSF loc*HSF/ slice=loc;
run;
```

b) Output:

```
RCBD WITH NESTED BLOCKING STRUCTURE
AND SUBSAMPLING
[t=3, b=8 (a=2,c=4), n=2]

The GLM Procedure

Class Level Information

Class          Levels    Values
loc              2      1 2
block            4      1 2 3 4
HSF              3      1 2 3

Number of Observations Read      48
Number of Observations Used      48
```

Table 9.16 (Continued)

The GLM Procedure					
Dependent Variable: height					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	40781.66667	1773.11594	80.44	<.0001
Error	24	529.00000	22.04167		
Corrected Total	47	41310.66667			
R-Square	Coeff Var	Root MSE	height Mean		
0.987195	2.228571	4.694855	210.6667		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
loc	1	20336.33333	20336.33333	922.63	<.0001
block(loc)	6	1462.33333	243.72222	11.06	<.0001
HSF	2	12170.66667	6085.33333	276.08	<.0001
loc*HSF	2	6511.16667	3255.58333	147.70	<.0001
block*HSF(loc)	12	301.16667	25.09722	1.14	0.3769
Source	DF	Type III SS	Mean Square	F Value	Pr > F
loc	1	20336.33333	20336.33333	922.63	<.0001
block(loc)	6	1462.33333	243.72222	11.06	<.0001
HSF	2	12170.66667	6085.33333	276.08	<.0001
loc*HSF	2	6511.16667	3255.58333	147.70	<.0001
block*HSF(loc)	12	301.16667	25.09722	1.14	0.3769
Tests of Hypotheses Using the Type III MS for block*HSF(loc) as an Error Term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
HSF	2	12170.66667	6085.33333	242.47	<.0001
loc*HSF	2	6511.16667	3255.58333	129.72	<.0001
The Mixed Procedure					
Covariance Parameter Estimates					
Cov Parm	Estimate				
block*HSF(loc)	1.5278				
Residual	22.0417				

Table 9.16 (Continued)

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
loc	1	12	810.30	<.0001
block(loc)	6	12	9.71	0.0005
HSF	2	12	242.47	<.0001
loc*HSF	2	12	129.72	<.0001

Least Squares Means

Effect	loc	HSF	Estimate	Standard Error	DF	t Value	Pr > t
HSF		1	202.00	1.2524	12	161.29	<.0001
HSF		2	233.00	1.2524	12	186.04	<.0001
HSF		3	197.00	1.2524	12	157.29	<.0001
loc*HSF	1	1	220.88	1.7712	12	124.70	<.0001
loc*HSF	1	2	268.63	1.7712	12	151.66	<.0001
loc*HSF	1	3	204.25	1.7712	12	115.32	<.0001
loc*HSF	2	1	183.13	1.7712	12	103.39	<.0001
loc*HSF	2	2	197.38	1.7712	12	111.44	<.0001
loc*HSF	2	3	189.75	1.7712	12	107.13	<.0001

Tests of Effect Slices

Effect	loc	Num DF	Den DF	F Value	Pr > F
loc*HSF	1	2	12	355.98	<.0001
loc*HSF	2	2	12	16.21	0.0004

-
- (i) The ANOVA is given as outlined in Table 9.10. The N effects are significantly different ($P < .0001$) and there is also significant variety \times N interaction ($P < .0072$).
 - (ii) Since we have quantitative treatments here, a post-hoc analysis should be a trend analysis. For the overall N-levels we find a significant linear and quadratic trend ($P = .0006$ and $P < .0001$, respectively), and the N LS means indicate that level 3 provides the highest yield at 149.0.
 - (iii) However, since the var \times N interaction is significant, it seems appropriate to perform the trend analysis separately for each variety. The results show essentially the same trends as in (ii), but the linear trend for variety 2 is not significant ($P < .3325$). As a result, the highest yields occur for different levels of N, namely level 4 (at 150.0) for variety 1 and level 2 (at 159.0) for variety 2. \square

Table 9.17 Generalized Randomized Block Design

a) Input statements:

```
data fert;
input N var y @@;
datalines;
1 1 104 1 1 114 1 2 109 1 2 124
2 1 134 2 1 130 2 2 154 2 2 164
3 1 146 3 1 142 3 2 152 3 2 156
4 1 150 4 1 150 4 2 140 4 2 135
5 1 133 5 1 146 5 2 131 5 2 137
;
run;

proc glm data=fert;
class var N;
model y=var N var*N;
lsmeans N var*N/stderr;
contrast 'N-linear' N -2 -1 0 1 2;
contrast 'N-quad' N 2 -1 -2 -1 2;
contrast 'N-linear var1' N -2 -1 0 1 2 var*N -2 -1 0 1 2 0 0 0 0;
contrast 'N-quad var1' N 2 -1 -2 -1 2 var*N -2 -1 0 1 2 0 0 0 0;
contrast 'N-linear var2' N -2 -1 0 1 2 var*N 0 0 0 0 0 -2 -1 0 1 2;
contrast 'N-quad var2' N 2 -1 -2 -1 2 var*N 0 0 0 0 0 2 -1 -2 -1 2;
title1 'GENERALIZED RANDOMIZED BLOCK DESIGN';
title2 '(t=5, b=2 r=2)';
title3 'ANOVA WITH POST-HOC TREND ANALYSIS';
run;
```

b) Output:

GENERALIZED RANDOMIZED BLOCK DESIGN		
(t=5, b=2 r=2)		
ANOVA WITH POST-HOC TREND ANALYSIS		
The GLM Procedure		
Class Level Information		
Class	Levels	Values
var	2	1 2
N	5	1 2 3 4 5
Number of Observations Read		
		20
Number of Observations Used		
		20

Table 9.17 (Continued)

The GLM Procedure					
Dependent Variable: y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	4465.450000	496.161111	14.12	0.0001
Error	10	351.500000	35.150000		
Corrected Total	19	4816.950000			
	R-Square	Coeff Var	Root MSE	y Mean	
	0.927029	4.310246	5.928744	137.5500	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
var	1	140.450000	140.450000	4.00	0.0735
N	4	3393.700000	848.425000	24.14	<.0001
var*N	4	931.300000	232.825000	6.62	0.0072
Source	DF	Type III SS	Mean Square	F Value	Pr > F
var	1	140.450000	140.450000	4.00	0.0735
N	4	3393.700000	848.425000	24.14	<.0001
var*N	4	931.300000	232.825000	6.62	0.0072
Least Squares Means					
	N	y LSMEAN	Standard Error	Pr > t	
	1	112.750000	2.964372	<.0001	
	2	145.500000	2.964372	<.0001	
	3	149.000000	2.964372	<.0001	
	4	143.750000	2.964372	<.0001	
	5	136.750000	2.964372	<.0001	
var	N	y LSMEAN	Standard Error	Pr > t	
1	1	109.000000	4.192255	<.0001	
1	2	132.000000	4.192255	<.0001	
1	3	144.000000	4.192255	<.0001	
1	4	150.000000	4.192255	<.0001	
1	5	139.500000	4.192255	<.0001	
2	1	116.500000	4.192255	<.0001	
2	2	159.000000	4.192255	<.0001	
2	3	154.000000	4.192255	<.0001	
2	4	137.500000	4.192255	<.0001	
2	5	134.000000	4.192255	<.0001	

Table 9.17 (Continued)

Dependent Variable: y

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
N-linear	1	855.625000	855.625000	24.34	0.0006
N-quad	1	2225.160714	2225.160714	63.30	<.0001
N-linear var1	1	1248.200000	1248.200000	35.51	0.0001
N-quad var1	1	761.285714	761.285714	21.66	0.0009
N-linear var2	1	36.450000	36.450000	1.04	0.3325
N-quad var2	1	1530.321429	1530.321429	43.54	<.0001

EXAMPLE 9.21: This example is intended to illustrate some features of the analysis of incomplete block designs, in particular the balanced incomplete block design. We consider here the BIBD (3, 3, 2, 2; 1) (which is not particularly useful from a practical point of view) with the data given in Table 9.18a.

We comment below on the input statements given in Table 9.18a and the output contained in Table 9.18b:

- (i) As options in the model statement we include “inverse” and “solution”. The 7×7 inverse given in the output is, of course, a g-inverse of the coefficient matrix of the NE. It is obtained by imposing the conditions $\beta_3 = 0$, $\hat{\tau}_3 = 0$. This is also reflected in the vector of solutions obtained in this way. For example, we have $\hat{\tau}_1 = 16.0$, $\hat{\tau}_2 = 3.0$, $\hat{\tau}_3 = 0$, so that we can obtain

$$\hat{\tau}_1 - \hat{\tau}_2 = 16.0 - 3.0 = 13.0$$

$$\hat{\tau}_1 - \hat{\tau}_3 = 16.0 - 0 = 16.0$$

$$\hat{\tau}_2 - \hat{\tau}_3 = 3.0 - 0 = 3.0$$

as the estimates of treatment differences. We emphasize here that the solution vector can only be used to obtain estimates of estimable functions, in particular treatment contrasts.

- (ii) The variance of the estimate of a treatment contrast can be obtained from the g-inverse matrix (see Section 4.16.2) as follows: Consider the 3×3 sub-matrix corresponding to the treatment effects

$$\begin{pmatrix} 1.333 & 0.667 & 0 \\ 0.667 & 1.333 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

We then obtain, for example,

$$\begin{aligned} \widehat{\text{var}}(\hat{\tau}_1 - \hat{\tau}_2) &= (1.333 - 2 \cdot 0.667 + 1.333) \hat{\sigma}_e^2 \\ &= 1.333 \text{ MS}(E) \\ &= 1.333 \cdot 1.5 = 2 \end{aligned}$$

From this we obtain the standard error (se) as

$$\text{se}(\hat{\tau}_1 - \hat{\tau}_2) = \sqrt{2} = 1.414$$

which agrees with the value given in the output.

- (iii) Looking at all possible treatment differences we see also that they all have the same standard error (1.414), a property of the BIBD.
- (iv) In the LSmeans statement we have included the option “e”. The resulting “Coefficients for trt Least Square Means” tells us how the LS mean for a particular treatment is obtained from the solution vector (since this is no longer simply the treatment mean). We illustrate this with the following example for treatment 1:

$$\begin{aligned} \text{LS mean (T}_1\text{)} &= 1 \cdot \hat{\mu} + .333(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3) + 1 \cdot \hat{\tau}_1 \\ &= 19.5 + .333(-12.0 - 6.0 - 0) + 16.0 \\ &= 19.5 - 6.0 + 16.0 = 29.5 \end{aligned}$$

The g-inverse can then be used, as illustrated earlier, to obtain the standard error for the LS means. \square

EXAMPLE 9.22: This is a continuation of Example 9.21 to illustrate the use of the reduced normal equation (RNE) in analyzing data from an incomplete block design.

In Table 9.19a we give the input statements, and Table 9.19b contains the output. We comment on both briefly as follows:

- (i) The main feature in the input statement is the “absorb block” statement. This results in absorbing the equations for μ and $\beta_1, \beta_2, \beta_3$ into the equations for τ_1, τ_2, τ_3 to obtain the RNE.
- (ii) The C-matrix (see (9.101)) can be displayed by including the “xpx” statement, and the “inverse” statement displays a g-inverse \mathbf{C}^- , obtained by imposing the condition $\hat{\tau}_3 = 0$. Note that this g-inverse is different from the expression given in (9.105) which is obtained by imposing the condition $\hat{\tau}_1 + \hat{\tau}_2 + \hat{\tau}_3 = 0$. But for both choices of a g-inverse estimates for treatment contrasts will be identical.
- (iii) The result of the test for $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ is the same as that given in Table 9.18b.
- (iv) By using the “absorb” option we cannot ask for treatment LS means because for this we need the solutions $\hat{\mu}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ which are not available now. \square

Table 9.18 Balanced Incomplete Block Design

a) Input statements:

```
data BIBD;
input block trt y;
datalines;
1 1 24
1 2 10
2 1 29
2 3 14
3 2 23
3 3 19
;
run;

proc glm data=BIBD;
class block trt;
model y=block trt/ inverse solution;
lsmeans trt/stderr e;
estimate '1 vs 2' trt 1 -1 0;
estimate '1 vs 3' trt 1 0 -1;
estimate '1 vs 3' trt 0 1 -1;
title1 'BALANCED INCOMPLETE BLOCK DESIGN';
title2 '(t=3, b=3, k=2, r=2, lambda=1)';
title3 'ANALYSIS OF VARIANCE';
title4 'W/POST-HOC ANALYSIS';
run;
```

b) Output:

BALANCED INCOMPLETE BLOCK DESIGN		
(t=3, b=3, k=2, r=2, lambda=1)		
ANALYSIS OF VARIANCE		
W/POST-HOC ANALYSIS		
The GLM Procedure		
Class Level Information		
Class	Levels	Values
block	3	1 2 3
trt	3	1 2 3
Number of Observations Read		6
Number of Observations Used		6

Table 9.18 (Continued)

The GLM Procedure

X'X Generalized Inverse (g2)

	Intercept	block 1	block 2	block 3
Intercept	0.8333333333	-0.3333333333	-0.6666666667	0
block 1	-0.3333333333	1.3333333333	0.6666666667	0
block 2	-0.6666666667	0.6666666667	1.3333333333	0
block 3	0	0	0	0
trt 1	-0.3333333333	-0.6666666667	-0.3333333333	0
trt 2	-0.6666666667	-0.3333333333	0.3333333333	0
trt 3	0	0	0	0
y	19.5	-12	-6	0

X'X Generalized Inverse (g2)

	trt 1	trt 2	trt 3	y
Intercept	-0.3333333333	-0.6666666667	0	19.5
block 1	-0.6666666667	-0.3333333333	0	-12
block 2	-0.3333333333	0.3333333333	0	-6
block 3	0	0	0	0
trt 1	1.3333333333	0.6666666667	0	16
trt 2	0.6666666667	1.3333333333	0	3
trt 3	0	0	0	0
y	16	3	0	1.5

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	241.3333333	60.33333333	40.22	0.1176
Error	1	1.5000000	1.5000000		
Corrected Total	5	242.8333333			

R-Square	Coeff Var	Root MSE	y Mean
0.993823	6.175184	1.224745	19.83333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
block	2	24.3333333	12.1666667	8.11	0.2410
trt	2	217.0000000	108.5000000	72.33	0.0829

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	2	108.0000000	54.0000000	36.00	0.1170
trt	2	217.0000000	108.5000000	72.33	0.0829

Table 9.18 (Continued)

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		19.50000000 B	1.11803399	17.44	0.0365
block	1	-12.00000000 B	1.41421356	-8.49	0.0747
block	2	-6.00000000 B	1.41421356	-4.24	0.1474
block	3	0.00000000 B	.	.	.
trt	1	16.00000000 B	1.41421356	11.31	0.0561
trt	2	3.00000000 B	1.41421356	2.12	0.2804
trt	3	0.00000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Least Squares Means

Coefficients for trt Least Square Means

Effect		trt Level		
		1	2	3
Intercept		1	1	1
block	1	0.33333333	0.33333333	0.33333333
block	2	0.33333333	0.33333333	0.33333333
block	3	0.33333333	0.33333333	0.33333333
trt	1	1	0	0
trt	2	0	1	0
trt	3	0	0	1

		y LSMEAN	Standard Error	Pr > t
1		29.5000000	0.9574271	0.0207
2		16.5000000	0.9574271	0.0369
3		13.5000000	0.9574271	0.0451

Dependent Variable: y

Parameter	Estimate	Standard Error	t Value	Pr > t
1 vs 2	13.0000000	1.41421356	9.19	0.0690
1 vs 3	16.0000000	1.41421356	11.31	0.0561
2 vs 3	3.0000000	1.41421356	2.12	0.2804

Table 9.19 Balanced Incomplete Block Design (RNE)

a) Input statements:

```
data BIBD;
input block trt y;
datalines;
1 1 24
1 2 10
2 1 29
2 3 14
3 2 23
3 3 19
;
run;
proc glm data=BIBD;
absorb block;
model y=trt/xpx inverse solution;
title1 'BALANCED INCOMPLETE BLOCK DESIGN';
title2 '(t=3, b=3, k=2, r=2, lambda=1)';
title3 'USING REDUCED NORMAL EQUATIONS';
run;
```

b) Output:

```
BALANCED INCOMPLETE BLOCK DESIGN
(t=3, b=3, k=2, r=2, lambda=1)
USING REDUCED NORMAL EQUATIONS

The GLM Procedure

Class Level Information

Class          Levels    Values
trt              3      1 2 3

Number of Observations Read          6
Number of Observations Used          6

The X'X Matrix

trt 1          trt 2          trt 3          y
trt 1              1          -0.5          -0.5          14.5
trt 2          -0.5              1          -0.5          -5
trt 3          -0.5          -0.5              1          -9.5
y              14.5          -5          -9.5          218.5

X'X Generalized Inverse (g2)

trt 1          trt 2          trt 3          y
trt 1          1.333333333333          0.666666666667          0          16
trt 2          0.666666666667          1.333333333333          0          3
trt 3              0              0              0          0
y              16              3              0          1.5
```

Table 9.19 (Continued)

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	241.3333333	60.3333333	40.22	0.1176
Error	1	1.5000000	1.5000000		
Corrected Total	5	242.8333333			

R-Square	Coeff Var	Root MSE	y Mean
0.993823	6.175184	1.224745	19.83333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
block	2	24.3333333	12.1666667	8.11	0.2410
trt	2	217.0000000	108.5000000	72.33	0.0829

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	217.0000000	108.5000000	72.33	0.0829

Parameter		Estimate	Standard Error	t Value	Pr > t
trt	1	16.00000000 B	1.41421356	11.31	0.0561
trt	2	3.00000000 B	1.41421356	2.12	0.2804
trt	3	0.00000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.



9.11 EXERCISES

- 9.1 Consider the following data from an experiment testing the effects of 5 levels of application of potash on the Pressley strength index of cotton (John and Que-nouille, 1977)

	Treatments*				
	1	2	3	4	5
Block 1	7.62	8.14	7.76	7.17	7.46
Block 2	8.00	8.15	7.73	7.57	7.68
Block 3	7.93	7.87	7.74	7.80	7.21

*Pounds K_2O per acre, expressed as units.

- (i) Obtain the ANOVA table and test the hypothesis that there are no differences among the treatments.
 - (ii) Since the treatments are quantitative, rather than making comparisons between individual treatments, it is preferable to explore the response curve. Partition the treatment sum of squares into three components due to linear effects, quadratic effects, and remainder. Test for linear and quadratic effects.
 - (iii) Suppose the actual levels of the treatments are 36, 54, 72, 108, and 144 pounds K_2O per acre, respectively. Partition the treatment sum of squares into a component due to linear response and a component due to deviation from linear response. Test for linearity.
 - (iv) Find the relative efficiency of this design and interpret the result.
- 9.2 Consider an experiment with 5 treatments in a RCB design with 10 blocks. The partial ANOVA table is as follows:

Source	d.f.	SS	MS
Blocks		135	
Treatments		100	
Residual			
Total		307	

- (i) Complete the ANOVA table above.
- (ii) Give the test statistic for testing $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$.
- (iii) Suppose a preplanned comparison of treatments is that of comparing treatment 1 against the average of treatments 2 and 3. Give the estimated variance for the estimate of this treatment comparison.

- (iv) Suppose you have supplementary information in the form of a variate x . Performing an analysis of covariance you obtain (among other quantities)

$$\sum_{i,k} (y_{ik} - \bar{y}_{i.} - \bar{y}_{.k} + \bar{y}_{..})(x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x}_{..}) = -20$$

$$\sum_{i,k} (x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x}_{..})^2 = 10$$

Obtain the estimate of σ_{e*}^2 .

- (v) Under the analysis of covariance model what is the estimator for the contrast in (iii) and its estimated variance (use numerical values where possible based on the information provided in this problem).

9.3 A chemist wants to compare three treatments. The experimental material he plans to use comes from four different manufacturers. He expects systematic differences among the material from the different manufacturers. Moreover, he is interested in finding out whether differences between treatment effects depend on the manufacturer. There is sufficient experimental material from each manufacturer for 12 experimental units.

- (i) What experimental design should he use for his experiment?
 (ii) Suppose he comes to you with the following partial ANOVA table:

Source	d.f.	SS	MS
Manufacturers (M)			20
Treatments (T)			25
$M \times T$			5
Error			2.3
<hr/>			
Total			

Fill in the ANOVA table.

- (iii) Give the SAS statements (classes, model) for the ANOVA in (ii).
 (iv) The chemist claims that he has used four replications per treatment, per manufacturer. After some questioning you find out that the four “replications” are actually two replications and two duplicate measurements for each experimental unit. In this case what should the ANOVA really look like? Give sources of variation and d.f.
 (v) Give the SAS statements (classes, model) for the ANOVA in (iv).
 (vi) Unfortunately, the computer file with his original data was destroyed. In order to correct the situation and come up with a reasonable ANOVA table, you devise a small experiment which will allow you to obtain an estimate of the sampling (observational) error variance component. Suppose this

estimate equals 1. Assuming that this is a reasonable estimate for the actual experiment, that is, assuming that this value was obtained from the actual experiment, complete now the ANOVA table given in (iv) by filling in d.f., SSs and MSs.

- (vii) Based on the ANOVA in (vi), give the test statistic for testing the null hypothesis that there is no $M \times T$ interaction.
- (viii) Based on the ANOVA in (vi), give the standard error for a simple treatment comparison (that is, the differences between the estimates of two treatment effects).

9.4 In a study of reaction time under the influence of alcohol, age is thought to be another factor that could affect the time. Test subjects (individuals) were classified into three age groups: 20–39, 40–59, 60 and over. In each age group each treatment (0 oz., 1 oz., 2 oz.) was randomly assigned to 4 individuals. The following results were obtained (the reaction time is measured in seconds):

Age	Amount of Alcohol		
	0 oz.	1 oz.	2 oz.
20–39	.42	.49	.65
	.45	.47	.60
	.39	.46	.70
	.40	.51	.66
40–59	.51	.70	1.05
	.55	.69	1.10
	.53	.73	.98
	.50	.75	1.12
60 and Over	.60	.85	1.25
	.59	.79	1.20
	.58	.88	1.30
	.61	.90	1.27

- (i) What kind of experimental design has been used?
- (ii) Give the appropriate ANOVA table.
- (iii) Investigate whether the differences in reaction time for the different amounts of alcohol depend on the age group.
- (iv) Is there a difference in reaction time for 1 oz. and 2 oz.? Does it make sense to consider this question?

9.5 Suppose a researcher comes to you with a table of data, obtained from a block design, that looks as follows:

	Treatment		
	1	2	3
Block 1	xxx	xxx	xxx
Block 2	xxx	xxx	xxx
Block 3	xxx	xxx	xxx
Block 4	xxx	xxx	xxx

where each x represents one observation. She asks you to analyze the data.

- (i) How would you determine how to analyze the data?
 - (ii) Describe two experimental plans that could have given rise to this data set.
 - (iii) For each of the plans (designs) you have identified in (ii) give
 - (a) an associated linear model,
 - (b) the ANOVA table based on this model,
 - (c) the variance of a simple treatment comparison,
 - (d) the estimator for the variance in (c).
 - (iv) Describe how you would do iii(b) in SAS, including the test of no treatment differences.
- 9.6** In studying the effects of pollution, seedlings are usually exposed to specified pollutants for a certain period of time (say 6 hours) during the day for several weeks after which an evaluation is made. The seedlings are put in a pollution chamber which then receives the pollutant. Consider a specific case in which an investigator wants to compare 4 pollutants: P_0 = filtered air, $P_1 = O_3$, $P_2 = NO_2$, and $P_3 = O_3 + NO_2$ at specified levels of concentration. He has only limited resources.
- In particular, he has only 8 pollution chambers in each of which he can put 3 seedlings. He feels that this is not adequate. So he decides to use the same chambers for 6 hours during the day and for 6 hours during the night, that is, 48 seedlings are used for the experiment. He is sure that there will be systematic differences between the day and night results.
- (i) Describe an experimental plan for this experiment. What is the name for the experimental design that he is using?
 - (ii) The investigator wonders whether certain comparisons among the pollutants will depend on whether one uses the results from day or night exposure. What kind of "effect" is he talking about and how can he investigate that?
 - (iii) Give an appropriate linear model for analyzing data from this experiment.
 - (iv) Outline the ANOVA table (source of variation, d.f., $E(MS)$) and indicate what hypotheses can be tested and how. In particular, what useful hypotheses about the treatment effects can be tested given the particular pollutants used in this experiment?

- 9.7** A researcher comes to you with data from a block design. For comparing 4 treatments he has used 5 blocks each with 4 experimental units. He took 2 measurements on each experimental unit. He comes to you with the following ANOVA table:

Source	d.f.	SS
Blocks	4	$SS(B)$
Treatments	3	$SS(T)$
Error	32	$SS(E)$
Total	39	

- (i) What name would you give to the experimental design that he used.
 - (ii) Give a linear model for the data from this design and outline the ANOVA (source of variation, d.f.).
 - (iii) Explain why or why not your ANOVA table in (ii) agrees with the researcher's ANOVA table given above.
 - (iv) Give the formula for the theoretical variance for the comparison between two treatments using this design.
- 9.8** Suppose a researcher comes to you for advice about the analysis of an experiment she has conducted. She has used 5 treatments and she has 9 observations for each treatment. She shows you the following ANOVA from a computer printout:

Source	d.f.	SS	F -value	$Pr > F$
Treatments	4	100	8.33	.0001
Error	40	120		

The researcher wants to know from you whether the analysis is correct, and if so what it means. For each of the following scenarios answer the following questions:

- (i) What is the name of the design?
- (ii) What is the appropriate model?
- (iii) Based on the model in (ii) what is the corresponding ANOVA (include sources of variation, d.f., form of test for testing $H_0 : \tau_1 = \tau_2 = \dots = \tau_5$).
- (iv) Is the ANOVA from the print-out above appropriate for testing $H_0 : \tau_1 = \dots = \tau_5$?
- (v) What are the SAS statements for doing (iii) and (iv)?

Scenario I: 45 animals were used for the study and each treatment was assigned to 9 animals at random;

Scenario II: 45 animals were used but they came from 9 different litters of size 5 each and each treatment was assigned to one animal in each litter;

Scenario III: 15 animals were used, each treatment was assigned to 3 animals at random, and 3 observations were obtained from each animal.

9.9 Consider the following experimental data:

	Treatment				
	1	2	3	4	5
Block 1	5.2	7.3		6.8	10.0
Block 2	8.5	9.2	12.7	10.6	11.6
Block 3	4.1	6.3	9.3	7.1	9.8
Block 4	6.4	8.9	11.2	5.7	11.3

- (i) Do the exact analysis, obtaining the ANOVA table, LS means for treatments, and the estimated variance for simple treatment comparisons (use SAS PROC GLM).
 - (ii) Estimate the missing value and obtain the approximate ANOVA.
 - (iii) Using the analysis of covariance technique, obtain a general expression for the treatment LS means and the variance of differences of LS means.
 - (iv) Obtain numerical values for the expressions in (iii) using the data above.
 - (v) Compare the results for (i) and (iv).
- 9.10** Obtain an expression for the bias of $SS(T)$ when the estimate of the missing value is used as if it were the real observation. [Hint: Compare $SS(T)$ with the SS obtained from the analysis of covariance.]
- 9.11** Consider a RCBD with subsampling. Specifically, suppose $t = 3$ treatments, $b = 4$ blocks, and $n = 2$ observations per experimental unit.

- (i) Give a linear model for data from such an experiment.
- (ii) Outline the ANOVA table, giving sources of variation, d.f., $E(MS)$.
- (iii) Write SAS statements to carry out the ANOVA in (ii).

Now suppose we have supplementary information in the form of a covariate x for each experimental unit.

- (iv) Give a linear model for analyzing data from such an experiment.

- (v) Outline the corresponding ANOVA table, giving sources of variation and d.f.
- (vi) Give a SAS statement for carrying out the ANOVA in (v).

Using the following data set

	Treatment		
	1	2	3
Block 1	(5, 7) 2	(10, 9) 4	(15, 18) 5
Block 2	(10, 11) 4	(14, 13) 5	(20, 22) 3
Block 3	(12, 15) 7	(18, 20) 3	(24, 27) 5
Block 4	(20, 18) 6	(21, 24) 2	(35, 33) 8

(where the two numbers in parentheses represent the duplicate observations and the number underneath represents the covariate x) perform the two ANOVAs as explained in (iii) and (vi).

9.12 A researcher wants to do an experiment using mice as the experimental units. She comes to you for advice on how to set up the experiment. Here is the situation: She wants to compare the effects of feeding three levels of calcium, say 0, 1, 2 units, on certain bone measurements (which can be observed only after the mice have been sacrificed). She has available 4 litters of 6 mice each, each litter containing 3 females and 3 males. Each litter comes from a different breed of mice.

- (i) Give two experimental designs that might be suitable for this experiment. For each design
 - (a) state any assumptions that would have to be made.
 - (b) Give the associated linear model.
 - (c) Sketch the ANOVA (source of variation, d.f., $E(\text{MS})$).
 - (d) State what inferences can be made.
- (ii) Explain how you would investigate the question that the two sexes react differently to the calcium treatments.
- (iii) The experimenter plans to make duplicate measurements on each animal. For both designs [given in (i)] give the estimators for the linear and quadratic effect of calcium and their standard error.

CHAPTER 10

Latin Square Type Designs

10.1 INTRODUCTION AND MOTIVATION

To introduce a new and important type of blocking, using two or more blocking factors (intrinsic and/or nonspecific), let us consider the following example.

EXAMPLE 10.1: Suppose a manufacturer wants to investigate and compare different production processes for ceramic cookware. Extraneous sources of systematic variability are identified as (i) different batches of raw material and (ii) different ovens used for baking the product. The batches of material are obtained from different sources and possibly at different times. The ovens available in the manufacturing plant are of different makes and different ages. If the raw material were indeed uniform, we would use a RCBD with the ovens as blocks. And if, on the other hand, the ovens were all of the same type, we would use a RCBD with the batches as blocks. In the present situation, however, we would like to use both batches and ovens as blocks in order to eliminate systematic variability and hence reduce the experimental error.

Suppose we have r batches of experimental material, B_1, B_2, \dots, B_r say, and c ovens O_1, O_2, \dots, O_c say. One way to proceed then perhaps might be to divide each batch into c equal parts and “form” blocks of the type $(B_i O_j)$, that is, combining each batch with each oven. If we have t treatments (processes) we would then mold t pieces of cookware, that is, t EUs, for each block $(B_i O_j)$ and then bake each piece in the assigned oven at the assigned process (treatment). In this whole experiment we would then have $rc t$ EUs and the observations would be analyzed according to the RCBD analysis with rc blocks and t treatments (see Section 9.6.7 for crossed blocking factors).

□

The above procedure may be feasible for some experiments, but impossible for others since it may require too many EUs. In other cases such an arrangement may be physically impossible. An example of this is a field experiment where the experimental field exhibits a gradient in two (orthogonal) directions. This too is a situation where one may want to utilize simultaneously two blocking factors or, to use a common phrase,

block in two directions. Error-control designs other than the one described above have been developed for that purpose. Such designs are referred to as designs for *two-way elimination of heterogeneity*, or *row-column designs*. We shall refer to them as *Latin square type designs* because in many cases the Latin square design, a special form of row-column design developed by R. A. Fisher (1925, 1926), is the basic building block of such designs.

EXAMPLE 10.1 (*continued*): To complete the discussion of our example in the context of the preceding remarks we shall give appropriate error-control designs for three situations: (i) $t = 4, r = 4, c = 4$; (ii) $t = 4, r = 8, c = 4$; and (iii) $t = 4, r = 8, c = 5$. These designs using rc EUs are given in Table 10.1 where the processes (treatments) are designated by A, B, C, D . To understand the schematic representation of these designs, let us look at design (i): A piece of cookware from batch 1 is produced according to process D and then baked in oven 1, a piece of cookware from batch 2 is produced according to process C and also baked in oven 1, and so on. \square

How these designs were obtained will be discussed in the following sections, but the reader should have no difficulty recognizing that these designs have certain structures and combinatorial properties: In design (i) each treatment (Latin letters) occurs once in each row and once in each column; in design (ii) each treatment occurs once in each row and twice in each column (rows 1, 2, 4, 7 constitute in fact design (i) and rows 3, 5, 6, 8 constitute design (i) with permuted columns); and design (iii) is an augmentation of design (ii) with column 5 being formed by columns 1 and 4 of design (i). We shall show in the following sections that these combinatorial structures make it possible to obtain estimates of treatment effect contrasts, which after all is the objective of the experiment.

As in these introductory remarks we shall begin with the simplest design for *two-way elimination of heterogeneity*, the Latin square design, which is then used as the building block for more complex designs involving two or more blocking factors.

10.2 LATIN SQUARE DESIGN

10.2.1 Definition

The *Latin square design* represents, in some sense, the simplest form of a row-column design. It is used for comparing t treatments in t rows and t columns, where rows and columns represent the two blocking factors. Latin squares and their combinatorial properties have been attributed to Euler (1782). They were proposed as experimental designs by Fisher (1925, 1926), although De Palluel (1788) already utilized the idea of a 4×4 Latin square design for an agricultural experiment (see Street and Street, 1987, 1988).

Mathematically speaking, the Latin square of order t is an arrangement of t Latin letters in a square of t rows and t columns such that every Latin letter occurs once in each row and once in each column (see design (i) in Table 10.1), or more generally, the arrangement of t symbols in a $t \times t$ array such that each symbol occurs exactly once in each row and column. In the context of experimental design, the Latin letters are the treatments. Latin squares exist for every t . A reduced Latin square (or Latin

Table 10.1 Latin Square Type Designs

(i) $t = r = c = 4$:

	O_1	O_2	O_3	O_4
B_1	D	B	C	A
B_2	C	A	B	D
B_3	B	D	A	C
B_4	A	C	D	B

(ii) $t = c = 4, r = 8$:

	O_1	O_2	O_3	O_4
B_1	D	B	C	A
B_2	C	A	B	D
B_3	A	D	B	C
B_4	B	D	A	C
B_5	D	C	A	B
B_6	C	B	D	A
B_7	A	C	D	B
B_8	B	A	C	D

(iii) $t = 4, r = 8, c = 5$:

	O_1	O_2	O_3	O_4	O_5
B_1	D	B	C	A	D
B_2	C	A	B	D	C
B_3	A	D	B	C	B
B_4	B	D	A	C	A
B_5	D	C	A	B	A
B_6	C	B	D	A	D
B_7	A	C	D	B	C
B_8	B	A	C	D	B

square in standard form) is one in which the first row and the first column are arranged in alphabetical order, for example, for $t = 3$,

<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>C</i>	<i>A</i>
<i>C</i>	<i>A</i>	<i>B</i>

this is the only reduced Latin square. The number of squares that can be generated from a reduced Latin square by permutation of the rows, columns, and letters is $(t!)^3$. These are not necessarily all different. If all rows but the first and all columns are permuted, we generate $t!(t - 1)!$ different squares. From the reduced Latin square of order 3 we can thus generate 12 squares.

In general, if the number of reduced squares of order t is denoted by T_t and the total number of Latin squares of order t by U_t , then $U_t = t!(t - 1)!T_t$ (see Dénes and Keedwell, 1974). Below we give a list of t, T_t for $t = 2, 3, \dots, 8$ (see e.g., Dénes and Keedwell, 1974):

t	T_t
2	1
3	1
4	4
5	56
6	9,408
7	16,942,080
8	535,281,401,856

10.2.2 Transformation Sets and Randomization

An enumeration of all possible reduced Latin squares is facilitated through the notion of *transformation sets* which is defined as follows: One square of the transformation set may be obtained from the others by permutation of letters and subsequent rearrangement into reduced or standard form. For $t = 4$ there exist two transformation sets as given below.

Set 1:

(1)	(2)	(3)
<i>A B C D</i>	<i>A B C D</i>	<i>A B C D</i>
<i>B A D C</i>	<i>B C D A</i>	<i>B D A C</i>
<i>C D B A</i>	<i>C D A B</i>	<i>C A D B</i>
<i>D C A B</i>	<i>D A B C</i>	<i>D C B A</i>

Set 2:

<i>A B C D</i>
<i>B A D C</i>
<i>C D A B</i>
<i>D C B A</i>

In the first transformation set, square (2) can be obtained from square (1) by interchanging A and D in (1) and rearranging, thus:

$$\begin{array}{cccc} D & B & C & A \\ B & D & A & C \\ C & A & B & D \\ A & C & D & B \end{array} \rightarrow \begin{array}{cccc} A & B & C & D \\ C & D & A & B \\ D & A & B & C \\ B & C & D & A \end{array} \rightarrow \begin{array}{cccc} A & B & C & D \\ B & C & D & A \\ C & D & A & B \\ D & A & B & C \end{array}$$

On the other hand, in set 2 any interchange of letters and rearrangement into standard form leads to its reproduction. Thus the two transformation sets above account for all $T_4 = 4$ reduced squares.

For $t = 5$ there exist two transformation sets, one containing 50 reduced squares and the other containing 6 reduced squares, thus accounting for all $T_5 = 56$ reduced squares.

For $t = 6$ there exist 22 transformation sets which contain a total of $T_6 = 9,408$ reduced Latin squares.

Fisher and Yates (1957) list examples of Latin squares for $t = 4, 5, \dots, 12$.

The actual randomization procedure for Latin square designs was given by Yates (1933) and is described by Fisher and Yates (1957) as follows. The first step is to select a reduced square at random. For squares of order 3, 4, or 5, the second step is to permute all rows except the first and all columns, or all rows and all columns except the first, and assign treatments at random to the letters A, B, C, \dots . For squares of order 6 permute all rows and columns, and then assign the letters to treatments at random. For larger squares, it is satisfactory to take any square and permute rows, columns, and treatments.

The randomization procedure for a 4×4 Latin square using SAS PROC PLAN is illustrated in Table 10.2, where the treatments are represented by the numbers 1, 2, 3, 4.

10.2.3 Derived Linear Model

We shall now examine the Latin square design (LSD) as an error-control design from the same point of view as that which we used for the RCBD. We suppose then that the subscripts (i, j, k) denote the row, column, and treatment of a particular EU. In all there are t^3 possible responses, for each treatment can conceptually be applied to each EU, and from this population of true yields we draw a sample which is based on a random $t \times t$ Latin square as described above. Such a sample has obvious properties of balance, particularly when we are concerned with the comparison of treatments.

Assuming unit-treatment additivity in the strict sense and following Kempthorne (1952) we write the conceptual response of treatment k in row i and column j as

$$T_{ijk} = U_{ij} + T_k, \quad (10.1)$$

where U_{ij} is the contribution from the EU in the i th row and j th column and T_k is the contribution from treatment k . We rewrite (10.1) as

$$\begin{aligned} T_{ijk} &= \bar{U}_{..} + (\bar{U}_{i.} - \bar{U}_{..}) + (\bar{U}_{.j} - \bar{U}_{..}) \\ &\quad + (U_{ij} - \bar{U}_{i.} - \bar{U}_{.j} + \bar{U}_{..}) + \bar{T}_{.} + (T_k - \bar{T}_{.}) \\ &= \mu + \rho_i + \gamma_j + \tau_k + u_{ij}, \end{aligned} \quad (10.2)$$

Table 10.2 Randomization for 4×4 Latin Square Design

a) Input statements:

```
factors rows=4 ordered cols=4 ordered/noprint;
treatments tmts=4 cyclic;
output out=LS
    rows cvals= ('B1' 'B2' 'B3' 'B4') random
    cols cvals= ('O1' 'O2' 'O3' 'O4') random
    tmts nvals= (1 2 3 4) random;
quit;
proc tabulate data=LS;
class rows cols;
var tmts;
table rows, cols*(tmts*f=6.) / rts=8;
title 'RANDOMIZATION FOR 4X4 LATIN SQUARE DESIGN'; run;
run;
```

(b) Output:

RANDOMIZATION FOR 4X4 LATIN SQUARE DESIGN

	cols			
	01	02	03	04
	tmts	tmts	tmts	tmts
	Sum	Sum	Sum	Sum
rows				
B1	1	2	4	3
B2	3	1	2	4
B3	2	4	3	1
B4	4	3	1	2

where

$$\mu = \bar{U}_{..} + \bar{T}.$$

is the average conceptual response,

$$\rho_i = \bar{U}_{i.} - \bar{U}_{..}$$

is defined as the i th row effect, with $\sum \rho_i = 0$,

$$\gamma_j = \bar{U}_{.j} - \bar{U}_{..}$$

is defined as the j th column effect, with $\sum \gamma_j = 0$,

$$\tau_k = T_k - \bar{T}.$$

is defined as the k th treatment effect, with $\sum \tau_k = 0$, and

$$\begin{aligned} u_{ij} &= U_{ij} - \bar{U}_{i.} - \bar{U}_{.j} + \bar{U}_{..} \\ &= (U_{ij} - \bar{U}_{..}) - (\bar{U}_{i.} - \bar{U}_{..}) - (\bar{U}_{.j} - \bar{U}_{..}) \end{aligned} \quad (10.3)$$

expresses the heterogeneity of the EUs in the sense that the contribution of EU (ij) is not made up additively of a row effect (ρ_i) and a column effect (γ_i). In fact, technically we may refer to u_{ij} as a row-column interaction effect.

We denote an actually observed response in the i th row and j th column by z_{ij} . Let δ_{ij}^k be the design random variable which takes the value unity if treatment k occurs on EU (i, j) and is zero otherwise. We can then write

$$z_{ij} = \sum_k \delta_{ij}^k T_{ijk}. \quad (10.4)$$

Substituting from (10.2) we obtain

$$z_{ij} = \mu + \rho_i + \gamma_j + \sum_k \delta_{ij}^k \tau_k + u_{ij}. \quad (10.5)$$

Alternatively, we can express an observation on a given EU in terms of the treatment it received. Let x_{kl} denote the observation from the l th application of treatment k and let ζ_{ij}^{kl} denote a random variable with $\zeta_{ij}^{kl} = 1$ if the l th application of treatment k falls on EU (i, j) and $\zeta_{ij}^{kl} = 0$, otherwise. Then

$$x_{kl} = \sum_{ij} \zeta_{ij}^{kl} z_{ij}. \quad (10.6)$$

We are mainly interested in treatment means and contrasts among them. From (10.6) we then have

$$\begin{aligned} \bar{x}_k &= \frac{1}{t} \sum_l \sum_{ij} \zeta_{ij}^{kl} z_{ij} \\ &= \frac{1}{t} \sum_{ij} \delta_{ij}^k z_{ij} \\ &= \mu + \tau_k + \frac{1}{t} \sum_{ij} \delta_{ij}^k u_{ij} \end{aligned} \quad (10.7)$$

since $\sum_l \zeta_{ij}^{kl} = \delta_{ij}^k$ and $\sum_{ij} \delta_{ij}^k = t$.

The joint distribution of the δ_{ij}^k 's is determined by the particular family of Latin squares from which the one actually used is chosen at random. If we use the randomization procedure described earlier we have

$$P(\delta_{ij}^k = 1) = \frac{1}{t}$$

and

$$P(\delta_{ij}^k = 0) = 1 - \frac{1}{t}.$$

The Latin square structure implies that if $\delta_{ij}^k = 1$ then $\delta_{i'j}^k = 0$ and $\delta_{ij'}^k = 0$. This implies

$$\begin{aligned} P(\delta_{ij}^k = 1, \delta_{i'j}^k = 1) &= 0 & (i \neq i') \\ P(\delta_{ij}^k = 1, \delta_{ij'}^k = 1) &= 0 & (j \neq j'). \end{aligned}$$

We also have

$$P(\delta_{ij}^k = 1, \delta_{i'j'}^k = 1) = \frac{1}{t(t-1)} \quad (i \neq i', j \neq j').$$

Having established some properties of the δ_{ij}^k 's we can now investigate the distributional properties of the models (10.5) and (10.6) or, more importantly, (10.7). Since

$$E_R(\delta_{ij}^k) = \frac{1}{t}$$

it follows from (10.7) immediately that

$$E_R(\bar{x}_{k.}) = \mu + \tau_k + \frac{1}{t^2} \sum_{ij} u_{ij}$$

or, using (10.3) and the fact that $\sum_i u_{ij} = \sum_j u_{ij} = 0$,

$$E_R(\bar{x}_{k.}) = \mu + \tau_k. \quad (10.8)$$

10.2.4 Estimation of Treatment Contrasts

It follows from (10.8) that a contrast among the treatment effects, $\sum_k c_k \tau_k$, with $\sum_k c_k = 0$, is estimated unbiasedly by the corresponding contrast among the treatment means, that is,

$$E_R \left(\sum_k c_k \bar{x}_{k.} \right) = \sum_k c_k \tau_k. \quad (10.9)$$

In order to obtain $\text{var}_R(\sum_k c_k \bar{x}_k.)$ we consider first

$$\begin{aligned}
 \text{var}_R(\bar{x}_k.) &= E_R[\bar{x}_k. - E_R(\bar{x}_k.)]^2 \\
 &= E_R \left[\frac{1}{t} \sum_{ij} \delta_{ij}^k u_{ij} \right]^2 \\
 &= \frac{1}{t^2} E_R \left[\sum_{ij} (\delta_{ij}^k)^2 u_{ij}^2 + \sum_{i \neq i'} \sum_j \delta_{ij}^k \delta_{i'j}^k u_{ij} u_{i'j} \right. \\
 &\quad \left. + \sum_i \sum_{j \neq j'} \delta_{ij}^k \delta_{ij'}^k u_{ij} u_{ij'} + \sum_{i \neq i'} \sum_{j \neq j'} \delta_{ij}^k \delta_{i'j'}^k u_{ij} u_{i'j'} \right] \\
 &= \frac{1}{t^2} \left[\frac{1}{t} \sum_{ij} u_{ij}^2 + 0 + 0 + \frac{1}{t(t-1)} \sum_{i \neq i'} \sum_{j \neq j'} u_{ij} u_{i'j'} \right] \tag{10.10}
 \end{aligned}$$

using the properties of the δ_{ij}^k 's. Now

$$\begin{aligned}
 \sum_{i \neq i'} \sum_{j \neq j'} u_{ij} u_{i'j'} &= \left(\sum_{ij} u_{ij} \right)^2 - \sum_{ij} u_{ij}^2 - \sum_{i \neq i'} \sum_j u_{ij} u_{i'j} - \sum_i \sum_{j \neq j'} u_{ij} u_{ij'} \\
 &= 0 - \sum_{ij} u_{ij}^2 + \sum_{ij} u_{ij}^2 + \sum_{ij} u_{ij}^2
 \end{aligned}$$

using again (10.3). We then have, substituting into (10.10),

$$\begin{aligned}
 \text{var}_R(\bar{x}_k.) &= \frac{1}{t^2} \left(\frac{1}{t} + \frac{1}{t(t-1)} \right) \sum_{ij} u_{ij}^2 \\
 &= \frac{1}{t^2(t-1)} \sum_{ij} u_{ij}^2. \tag{10.11}
 \end{aligned}$$

Similarly, we can obtain for $k' \neq k$

$$\text{cov}_R(\bar{x}_k., \bar{x}_{k'}.) = -\frac{1}{t^2(t-1)^2} \sum_{ij} u_{ij}^2. \tag{10.12}$$

Using (10.11) and (10.12) it then follows that

$$\begin{aligned}
 \text{var}_R \left(\sum_k c_k \bar{x}_k. \right) &= \frac{1}{t^2} \left[\sum_k c_k^2 \frac{1}{t-1} \sum_{ij} u_{ij}^2 - \sum_{k \neq k'} c_k c_{k'} \frac{1}{(t-1)^2} \sum_{ij} u_{ij}^2 \right] \\
 &= \sum_k c_k^2 \frac{1}{t(t-1)^2} \sum_{ij} u_{ij}^2. \tag{10.13}
 \end{aligned}$$

If we define

$$\sigma_u^2 = \frac{1}{(t-1)^2} \sum_{ij} u_{ij}^2 \quad (10.14)$$

we can write (10.13) as

$$\text{var}_R \left(\sum_k c_k \bar{x}_k \right) = \sum_k c_k^2 \frac{\sigma_u^2}{t}. \quad (10.15)$$

The problem then remains to estimate σ_u^2 . This is achieved again by means of the analysis of variance.

10.2.5 Analysis of Variance

To write out the ANOVA table it is convenient to write the observation on EU (i, j) which has received treatment k as $y_{ij(k)}$ where (\cdot) indicates that not all possible triplets (i, j, k) occur in the array, only t^2 out of t^3 . With this notation we then have

$$\begin{aligned} \bar{y}_{ij(\cdot)} &= z_{ij} \\ \bar{y}_{i(\cdot)} &= \bar{z}_{i\cdot}, \bar{y}_{j(\cdot)} = \bar{z}_{\cdot j} \\ \bar{y}_{\cdot(k)} &= \bar{x}_k. \\ \bar{y}_{\cdot(\cdot)} &= \bar{z}_{\cdot\cdot} = \bar{x}_{\cdot\cdot}. \end{aligned}$$

The ANOVA as given in Table 10.3 is based on the identity for $y_{ij(k)}$:

$$\begin{aligned} y_{ij(k)} &= \bar{y}_{\cdot(\cdot)} + (\bar{y}_{i(\cdot)} - \bar{y}_{\cdot(\cdot)}) + (\bar{y}_{j(\cdot)} - \bar{y}_{\cdot(\cdot)}) \\ &\quad + (\bar{y}_{\cdot(k)} - \bar{y}_{\cdot(\cdot)}) + (y_{ij(k)} - \bar{y}_{i(\cdot)} - \bar{y}_{j(\cdot)} - \bar{y}_{\cdot(k)} + 2\bar{y}_{\cdot(\cdot)}) \end{aligned}$$

and is obtained by squaring both sides and summing over all occurring combinations $ij(k)$.

The $E(\text{MS})$ given in the left-hand column are based on the models (10.5) and (10.6) as established earlier and follow from randomization theory. The results show that the LSD is an unbiased design in Yates' sense and that $\text{MS}(E)$ is an estimate of σ_u^2 .

To test the hypothesis $H_0 : \tau_1 = \tau_2 = \cdots = \tau_t = 0$ we are led by the $E(\text{MS})$ in Table 10.3 to consider the ratio

$$F = \frac{\text{MS}(T)}{\text{MS}(E)}$$

This ratio F will be evaluated for the square actually used and for all other squares that we could have obtained by the randomization procedure. If the value for the square actually used is equaled or exceeded by that of P percent of the possible arrangements (including the one used), we shall say that we have significance at the P percent level. The evaluation of the significance in a particular experiment could be somewhat laborious, and we rely on the fact that, similar to the result for the RCBD (see Section 9.2.5), the distribution of the criterion F will be closely approximated by the F -distribution with $t-1$ and $(t-1)(t-2)$ d.f.

Table 10.3 ANOVA for LSD

Source	d.f.	SS	MS	E(MS)	
				Add. in strict sense	Add. in broad sense
Rows	$t - 1$	$t \sum_i (\bar{y}_{i.(\cdot)} - \bar{y}_{..(\cdot)})^2 = SS(R)$	$MS(R) = SS(R)/(t - 1)$	$t \sum_i \rho_i^2/(t - 1)$	$\sigma_\eta^2 + \sigma_\nu^2 + t \sum_i \rho_i^2(t - 1)$
Columns	$t - 1$	$t \sum_j (\bar{y}_{.j(\cdot)} - \bar{y}_{..(\cdot)})^2 = SS(C)$	$MS(C) = SS(C)/(t - 1)$	$t \sum_j \gamma_j^2/(t - 1)$	$\sigma_\eta^2 + \sigma_\nu^2 + t \sum_j \gamma_j^2(t - 1)$
Treatments	$t - 1$	$t \sum_k (\bar{y}_{..(k)} - \bar{y}_{..(\cdot)})^2 = SS(T)$	$MS(T) = SS(T)/(t - 1)$	$\sigma_u^2 + t \sum_k \tau_k^2/(t - 1)$	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2 + t \sum_k \tau_k^2(t - 1)$
Error	$(t - 1)(t - 2)$	By subtraction = $SS(E)$	$MS(E) = SS(E)/(t - 1)(t - 2)$	σ_u^2	$\sigma_\eta^2 + \sigma_\nu^2 + \sigma_u^2$
Total	$t^2 - 1$	$\sum_{i,j(k)} (y_{ij(k)} - \bar{y}_{..(\cdot)})^2$			

The extent to which the distribution of the criterion F over the possible randomizations may be represented by the F -distribution has been examined by Welch (1937). He showed that the quantity

$$U = \frac{SS(T)}{SS(T) + SS(E)}$$

has a mean value of $1/(t-1)$, and this is the mean value of the beta distribution which is a transform of the F -distribution. This was obtained with only the assumption that

$$P(\delta_{ij}^k = 1, \delta_{i'j'}^k = 1) = \frac{1}{t(t-1)}$$

for $i \neq i', j \neq j'$ (see Section 10.2.3) and therefore holds for any transformation set. Welch (1937) also found that $\text{var}(U)$ does depend on the transformation set, and also on the quantities

$$\begin{aligned} C &= \sum_{ij} u_{ij}^4, & D &= \left(\sum_{ij} u_{ij}^2 \right)^2 \\ G &= \sum_i \left(\sum_j u_{ij}^2 \right)^2 + \sum_j \left(\sum_i u_{ij}^2 \right)^2 \\ H &= \sum_{ii'} \left(\sum_j u_{ij} u_{i'j} \right)^2, \end{aligned}$$

where u_{ij} is as defined in (10.3). He examined $\text{var}(U)$ for some constructed data and for some sets of uniformity data and found the proportion of times the 5 percent value of U from the beta distribution was exceeded ranged from 2.7 to 6.2 percent. The approximation by the F -distribution is therefore not entirely satisfactory, but the evidence is not conclusive, in that the approximation depends on the quantities C, D, G, H above, and, in a particular case, we do not know these values, nor do we know the values we shall meet in practice. The rules given above for the choice of a random Latin square are designed to give equal probability to all possible Latin squares of size less than 7×7 , and this appears to be a desirable procedure. To conclude this aspect of the Latin square, we shall assume that normal theory gives satisfactory approximations to corresponding randomization tests, but some care needs to be taken with small Latin squares as noted below.

We consider first the 2×2 Latin square, for which there are only 2 different ones: namely,

$$\begin{array}{cc} \text{A} & \text{B} \\ \text{B} & \text{A} \end{array} \quad \text{and} \quad \begin{array}{cc} \text{B} & \text{A} \\ \text{A} & \text{B} \end{array}$$

This square has no degrees of freedom for error, as is obvious from the fact that, if we use one square and obtain the treatment difference, then the treatment difference given by the other square is the negative of the difference with the former square. If then we

wish to compare 2 treatments with 2×2 Latin squares, we must use many squares, and to make any test we must assume that the difference is constant from square to square (see Section 10.3). The randomization test is so simply performed in this case, that with a small number of squares, say 6 or less, it would probably be advisable to rely on the randomization test procedures rather than the usual F -distribution approximation.

In the case of the 3×3 Latin square, it is important to note that there are, in fact, only 2 different partitions of the 9 cells into 3 sets of 3, in such a way that each set is represented in each row and in each column. There are 12 different 3×3 Latin squares, but these give the same partitioning in sets of 3. If we wish to test the null hypothesis that there are no differences among the 3 treatments, we shall use the ratio of treatment mean square to error mean square as the test criterion. If this takes the value R with the randomization we in fact used, it will take the same value for 5 of the other 11 randomizations, and the value $1/R$ for the remaining 6. This happens because the sum of the treatment sum of squares with 2 degrees of freedom and the error sum of squares with 2 degrees of freedom is constant for all randomizations, and a randomization that gives a partitioning different from the one actually used will have the treatment and error sum of squares interchanged. We are, therefore, in the position of not being able to make a significance test, for which the chance of rejecting the hypothesis when true is less than 50 percent (or, in other words, of size less than .50). This fact is important because, if we use the normal theory model in which the errors in the model are assumed to be normally and independently distributed with mean zero and constant variance σ^2 , we shall use the F -test with 2 and 2 degrees of freedom and can make a test at any significance level we please. The distinction we make throughout this book between the derived and normal theory model is therefore extremely relevant. If we consider a particular treatment contrast, and evaluate it for the 12 possible 3×3 Latin squares, we shall find that there are 6 possible values which the criterion (mean square due to treatment comparison/error mean square) can take. We therefore only make a test of significance with level 1-in-6, of the hypothesis that the true comparison is zero, if we use a 2-tailed test. For these reasons, a single 3×3 Latin square experiment is virtually valueless, and, if we use a small number of replications (see Section 10.3), we should, as with the 2×2 square, probably use the randomization test procedures, although it is often found that the usual F -test gives a remarkably similar answer.

There are in all $4(4!3!)$ or 576 different 4×4 Latin squares, but these lead to only 24 different partitions of the 16 cells into 4 sets of 4, in such a way that each cell is represented in each row and in each column. It is therefore desirable to make the test strictly according to the randomization test procedure.

Squares of side 5 and 6 were examined by Welch (1937). For squares of side 7 or more it seems reasonable to assume that the F -distribution is satisfactory.

10.2.6 The Model under Additivity in the Broad Sense

So far we have considered only the situation where unit-treatment additivity in the strict sense holds. To do so is useful to bring out some of the essential features of a LSD. From a practical point of view, however, the inclusion of technical errors is important, which leads us to the situation where unit-treatment additivity in the broad sense holds. We shall not go into all the details here but rather refer the reader to

our discussion of this topic in connection with the CRD (Chapter 6) and the RCBD (Chapter 9). The important point to remember here is that to the unit error, u_{ij} , we now add another component of experimental error, $\nu_{ij(k)}$, and the observational error, $\eta_{ij(k)}$, both with means zero and variances σ_ν^2 and σ_η^2 , respectively. The $E(\text{MS})$ in terms of these variance components are given in the right-hand column of Table 10.3.

As we have argued earlier (Chapter 6 and 9), for purposes of making inference about the treatments, it is convenient to define the experimental error as $\varepsilon_{ij(k)}$ and the observational error as $\eta_{ij(k)}$, which can be considered as i.i.d. random variables with means zero and variances $\sigma_\varepsilon^2 = \sigma_u^2 + \sigma_\nu^2$ and σ_η^2 , respectively. Going one step further we define the total error as $e_{ij(k)}$ which can be considered as i.i.d. random variables with mean zero and variance $\sigma_e^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$. We then write (following (10.5)) the model for an observation on EU (ij) to which treatment k was applied as

$$y_{ij(k)} = \mu + \rho_i + \gamma_j + \tau_k + e_{ij(k)}. \quad (10.16)$$

Its properties follow from our discussion above. In particular we obtain, as suggested by (10.15), that the treatment contrast $\sum c_k \tau_k$ is estimated as $\sum c_k \bar{y}_{..(k)}$ with

$$\text{var} \left(\sum_k c_k \hat{\tau}_k \right) = \text{var} \left(\sum_k c_k \bar{y}_{..(k)} \right) = \sum_k c_k^2 \frac{\sigma_e^2}{t}.$$

As is obvious from Table 10.3, $\text{MS}(E)$ is an estimate of σ_e^2 , and $H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$ can be tested as before by considering $F = \text{MS}(T)/\text{MS}(E)$ as an F -statistic with $t - 1$ and $(t - 1)(t - 2)$ d.f.

We point out that similar to our findings for the RCBD, the form of the $E(\text{MS})$ in Table 10.3 indicates that there do not exist legitimate tests for row and column effects. For an assessment of the effectiveness of blocking by rows or columns we refer to Section 10.2.9(iii).

10.2.7 Consequences of Nonadditivity

Just as in the case of the RCBD the assumption of unit-treatment additivity may not always hold. Wilk and Kempthorne (1957) discussed the LSD in its most general form. They considered the situation where the t rows are sampled from a population of R rows, the t columns are sampled from a population of C columns, and the t treatments are sampled from a population of T treatments. They also amended model (10.16) to include row \times treatment, column \times treatment, and row \times column \times treatment interactions. If σ_{rt}^2 , σ_{ct}^2 , and σ_{rct}^2 denote the variance components due to these interactions (for a precise definition we refer the reader to Wilk and Kempthorne, 1957), then the relevant $E(\text{MS})$ from the ANOVA table can be written as follows:

$$E[\text{MS}(T)] = \sigma_e^2 + \left(\phi + \frac{t}{RC} \right) \sigma_{rct}^2 + \frac{R-t}{R} \sigma_{rt}^2 + \frac{C-t}{C} \sigma_{ct}^2 + t \sum \tau_k^2 / (T-1) \quad (10.17)$$

$$E[\text{MS}(E)] = \sigma_e^2 + \phi \sigma_{rct}^2 + \sigma_{rt}^2 + \sigma_{ct}^2, \quad (10.18)$$

where

$$\phi = \left(1 - \frac{1}{R} - \frac{1}{C} - \frac{1}{T}\right).$$

For the special case $R = C = T = t$, which we have considered here, (10.17) and (10.18) reduce to

$$E[\text{MS}(T)] = \sigma_e^2 + \left(1 - \frac{2}{t}\right) \sigma_{rct}^2 + t \sum_k \tau_k^2 / (t - 1) \quad (10.19)$$

and

$$E[\text{MS}(E)] = \sigma_e^2 + \left(1 - \frac{3}{t}\right) \sigma_{rct}^2 + \sigma_{rt}^2 + \sigma_{ct}^2. \quad (10.20)$$

The results (10.17) and (10.18) suggest that if $R \gg t$ and $C \gg t$ (corresponding essentially to the situation of random row and column effects) then the usual F -test suggested above is still appropriate even in the presence of interactions. This is, however, not true for the fixed effects situation as illustrated by comparing (10.19) and (10.20). In this case $\text{MS}(E)$ is on the average larger than $\text{MS}(T)$ under the hypothesis of no treatment effects and hence the usual F -test will lead to fewer significant results. In this case the LSD is not an unbiased design. For more details the reader is referred to Wilk and Kempthorne (1957) and an interesting somewhat different discussion by Neyman et al. (1935). Another objection to the assumption of additivity is provided by Srivastava and Wang (1998).

10.2.8 Investigating Nonadditivity

The problem of unit-treatment interaction, mainly in the form of row-treatment and/or column-treatment interaction, is obviously an important one. There is, however, no easy method of detecting such interactions in a LSD. A partial solution has been suggested by Tukey (1955). His method was reformulated in terms of an analysis of covariance by Rojas (1973) using an interaction model of the form

$$y_{ij(k)} = \mu + \rho_i + \gamma_j + \tau_k + \theta(\rho_i \tau_k + \gamma_j \tau_k + \rho_i \gamma_j) + e_{ij(k)}. \quad (10.21)$$

Writing (10.21) alternatively as

$$y_{ij(k)} = \mu + \rho_i + \gamma_j + \tau_k + \theta x_{ij(k)} + e_{ij(k)} \quad (10.22)$$

and choosing

$$\begin{aligned} x_{ij(k)} &= (\hat{\rho}_i + \hat{\gamma}_j + \hat{\tau}_k)^2 \\ &= 2(\hat{\rho}_i \hat{\tau}_k + \hat{\gamma}_j \hat{\tau}_k + \hat{\rho}_i \hat{\gamma}_j) + (\hat{\rho}_i^2 + \hat{\gamma}_j^2 + \hat{\tau}_k^2) \end{aligned} \quad (10.23)$$

Rojas showed that, as a generalization of the method described in Section 9.6, testing for $H_0 : \theta = 0$ with model (10.22) is the same as Tukey's one-degree-of-freedom test for nonadditivity in the LSD.

Model (10.21) is obviously only one of several ways in which interactions in the LSD can be characterized. Rather than including all two-factor interaction terms as in

model (10.21) it may be more useful to (i) include only treatment-row and treatment-column interactions and (ii) include them as separate terms in an analysis of covariance model, thus simply extending Scheffé's (1959) derivation of Tukey's (1949) test (see Section 9.6). We propose to model the interactions $(\rho\tau)_{ik}$ and $(\gamma\tau)_{jk}$ as $(\rho\tau)_{ik} = \rho_i\tau_k$ and $(\gamma\tau)_{jk} = \gamma_j\tau_k$ and consider the model

$$y_{ij(k)} = \mu + \rho_i + \gamma_j + \tau_k + \theta_1 x_{ij(k)} + \theta_2 z_{ij(k)} + e_{ij(k)}^* \quad (10.24)$$

with

$$x_{ij(k)} = \hat{\rho}_i \hat{\tau}_k = (\bar{y}_{i..} - \bar{y}_{..})(\bar{y}_{..(k)} - \bar{y}_{..})$$

for all j , and

$$z_{ij(k)} = \hat{\gamma}_j \hat{\tau}_k = (\bar{y}_{.j.} - \bar{y}_{..})(\bar{y}_{..(k)} - \bar{y}_{..})$$

for all i . Using the fact that $\bar{x}_{i..} = \bar{x}_{..(k)} = 0$, $\bar{z}_{.j.} = \bar{z}_{..(k)} = 0$ and applying the method described in Section 8.7 (modified for the LSD), we obtain immediately

$$\hat{\theta}_1 = \frac{E_{xy}}{E_{xx}} = \frac{\sum_{ijk} x_{ij(k)} y_{ij(k)}}{\sum_{ijk} x_{ij(k)}^2} = \frac{\sum_{ik} (\bar{y}_{i..} - \bar{y}_{..})(\bar{y}_{..(k)} - \bar{y}_{..}) \bar{y}_{i..(k)}}{\sum_i (\bar{y}_{i..} - \bar{y}_{..})^2 \sum_k (\bar{y}_{..(k)} - \bar{y}_{..})^2}$$

and

$$\hat{\theta}_2 = \frac{E_{zy}}{E_{zz}} = \frac{\sum_{ijk} z_{ij(k)} y_{ij(k)}}{\sum_{ijk} z_{ij(k)}^2} = \frac{\sum_{jk} (\bar{y}_{.j.} - \bar{y}_{..})(\bar{y}_{..(k)} - \bar{y}_{..}) \bar{y}_{.j.(k)}}{\sum_j (\bar{y}_{.j.} - \bar{y}_{..})^2 \sum_k (\bar{y}_{..(k)} - \bar{y}_{..})^2}$$

(we have used the fact here that $E_{xz} = 0$). We then obtain in the usual way

$$SS(\theta_1) = \frac{E_{xy}^2}{E_{xx}} \quad \text{and} \quad SS(\theta_2) = \frac{E_{zy}^2}{E_{zz}}.$$

To test $H_{0i} : \theta_i = 0$ ($i = 1, 2$) we use

$$F_i = \frac{SS(\theta_i)}{[SS(E) - SS(\theta_1) - SS(\theta_2)] / [(t-1)(t-2) - 2]}$$

as an F -statistic with 1 and $(t-1)(t-2) - 2$ d.f., where $SS(E)$ is obtained from Table 10.3.

As an alternative to model (10.24) we may, of course, only want to include one interaction term, depending on our knowledge of the experimental situation at hand. The modification to the F -test given above is obvious. Whichever model we consider, however, if interaction is indicated there does not seem to be an easy solution to a meaningful analysis of the data. Search for a suitable transformation to additivity may be an option, but it may not be easily achievable.

10.2.9 Miscellaneous Remarks

We conclude this section with a few remarks, concerning topics that were discussed in great detail in earlier chapters:

- (i) *Analysis of covariance.* In addition to the blocking by rows and columns supplementary information may be available on the EUs and further reduction in experimental error variance may be achieved by using analysis of covariance. The formal procedure is similar to that discussed in Section 9.4 with the obvious and by now familiar modifications.
- (ii) *Missing observations.* The method proposed by Coons (1957) as discussed in Section 9.5 can be applied here also. The formula for a missing value in row i^* , column j^* , and treatment k^* , corresponding to the development in Section 9.5.1, is now given by

$$\hat{\gamma} = \frac{tR_{i^*} + tC_{j^*} + tT_{k^*} - 2G}{(t-1)(t-2)} \quad (10.25)$$

using obvious notation. Formulae for several missing values can be obtained using several covariates and the methods of Section 8.7. Explicit expressions are given by Kramer and Glass (1960).

- (iii) *Relative efficiency.* It may be of interest to ask whether blocking in two directions has been useful compared to blocking in only one direction, either by rows or by columns, that is, using a RCBD with either rows as blocks or columns as blocks. Using the concept of a uniformity trial (see Section 9.3) and the resulting partition in the ANOVA, or randomization analysis,

Source	d.f.
Rows	$t - 1$
Columns	$t - 1$
Residual	$(t - 1)^2$

we obtain the following EREs:

- (a) Rows used as blocks

$$\text{ERE (LSD to RCBD}_{\text{Rows}}) = \frac{\text{MS}(C) + (t-1)\text{MS}(E)}{t\text{MS}(E)} \quad (10.26)$$

- (b) Columns used as blocks

$$\text{ERE (LSD to RCBD}_{\text{Columns}}) = \frac{\text{MS}(R) + (t-1)\text{MS}(E)}{t\text{MS}(E)} \quad (10.27)$$

where $\text{MS}(R)$, $\text{MS}(C)$, and $\text{MS}(E)$ are obtained from the ANOVA (see Table 10.3) of the completed experiment using an LSD. For another connection between the EREs and the ANOVA table see Lentner, Arnold, and Hinkelmann (1989).

It is sometimes argued that the LSD is not very useful from a practical point of view because of two limitations: (i) The numbers of rows, columns and treatments have to be the same, and (ii) for small values of t one has insufficient d.f. for error. We do not necessarily agree with such a strong viewpoint, but even critics agree that what we might call the *Latin square principle*, namely orthogonal blocking in two directions, is extremely important in the whole endeavor of experimental design. It gives rise to many more error-reduction designs as well as treatment designs which are of great practical value. Some such designs will be discussed in the next sections (see also Section II.6.6).

10.3 REPLICATED LATIN SQUARES

10.3.1 Different Scenarios for Replication

One way to increase the error d.f. is obviously to replicate a LSD. How such replications are carried out depends on the particular experimental situation. We shall use the example of Section 10.1 and illustrate, in accordance with our discussion in Chapter 2 and Section 9.6.7, how different methods of replication lead to different linear models and hence to different analyses.

Referring to Example 10.1, consider the following situations:

EXAMPLE 10.2: The basic experiment, using a LSD of order t , is replicated by the manufacturer r times as follows. Each of the t batches of experimental material is divided into r parts to be used in the r replications, respectively. The same t ovens are used in each replication. \square

EXAMPLE 10.3: Different batches of experimental material are obtained for each replication, t batches for each replication. The same t ovens are used in each replication. \square

EXAMPLE 10.4: Rather than having the experiment replicated by one manufacturer, we may ask r different manufacturers to carry out the basic experiment. Each manufacturer has his own suppliers of raw material and different ovens are available in the r different factories. \square

In each of these experiments the randomization procedure, as described in Section 10.2, is carried out independently for each component Latin square. The major difference between the situations described above is whether, in classification terminology, the various blocking factors are crossed with each other or nested within each other (see Section 4.12). We shall discuss this now and show how that affects the analysis.

Table 10.4 ANOVA for Model (10.28)

Source	d.f.	SS	$E(MS)$
Replications	$r - 1$	$t^2 \sum_i (\bar{y}_{i..(\cdot)} - \bar{y}_{...(\cdot)})^2$	
Batches	$t - 1$	$rt \sum_j (\bar{y}_{.j.(\cdot)} - \bar{y}_{...(\cdot)})^2$	
Ovens	$t - 1$	$rt \sum_k (\bar{y}_{..k(\cdot)} - \bar{y}_{...(\cdot)})^2$	
Treatments	$t - 1$	$rt \sum_l (\bar{y}_{...l}) - \bar{y}_{...(\cdot)})^2$	$\sigma_e^2 + rt \sum_l \tau_l^2 / (t - 1)$
Error	$r(t - 1)(t - 2) + 3(r - 1)(t - 1)$	Difference	σ_e^2
Total	$rt^2 - 1$	$\sum_{ijk(l)} (y_{ijk(l)} - \bar{y}_{...(\cdot)})^2$	

10.3.2 Rows and Columns Crossed with Replications

In Example 10.2 the factor “batches” and the factor “ovens” are crossed with the factor “replications” since the batches and ovens are the same in each replication. We assume that there are systematic differences between replications simply because of the time lag. As an extension of model (10.16) an appropriate model for observations from this design is then

$$y_{ijk(l)} = \mu + \alpha_i + \rho_j + \gamma_k + \tau_l + e_{ijk(l)}, \quad (10.28)$$

where ρ_j, γ_k, τ_l are as defined before and $\alpha_i (i = 1, 2, \dots, r)$ represents the effect of the i th replicate. This model leads to the ANOVA of Table 10.4.

We should point out that (10.28) is only one possible model. If desired and warranted this experimental setup allows us to separate out from the $SS(\text{Error})$ a sum of squares due to replication \times treatment interaction with $(r - 1)(t - 1)$ d.f. In fact, technically, the d.f. for error as given in Table 10.4 are the sum of the error d.f. for the r individual Latin squares and the d.f. for the replication \times row, replication \times column, and replication \times treatment interactions.

10.3.3 Rows Nested in and Columns Crossed with Replications

Since in Example 10.3 new batches of material are obtained for each replication, the factor “batches” is nested within the factor “replications.” As before, the factors “ovens” and “replications” are crossed. This leads to a model of the form

$$y_{ijk(l)} = \mu + \alpha_i + \rho_{ij} + \gamma_k + \tau_l + e_{ijk(l)}, \quad (10.29)$$

where the $\rho_{ij} (i = 1, 2, \dots, r; j = 1, 2, \dots, t)$ are now the row effects nested within replications. The ANOVA associated with model (10.29) is given in Table 10.5. Just as before, model (10.29) can be amended to include replication \times treatment interaction.

Table 10.5 ANOVA for Model (10.29)

Source	d.f.	SS	$E(MS)$
Replications	$r - 1$	$t^2 \sum_i (\bar{y}_{i..(\cdot)} - \bar{y}_{...(\cdot)})^2$	
Batches/Reps	$r(t - 1)$	$t \sum_{i,j} (\bar{y}_{ij.(\cdot)} - \bar{y}_{i..(\cdot)})^2$	
Ovens	$t - 1$	$rt \sum_k (\bar{y}_{..k(\cdot)} - \bar{y}_{...(\cdot)})^2$	
Treatments	$t - 1$	$rt \sum_l (\bar{y}_{...l}) - \bar{y}_{...(\cdot)})^2$	$\sigma_e^2 + rt \sum_l \tau_l^2 / (t - 1)$
Error	$r(t - 1)(t - 2) + 2(r - 1)(t - 1)$	Difference	σ_e^2
Total	$rt^2 - 1$	$\sum_{ijk(l)} (y_{ijk(l)} - \bar{y}_{...(\cdot)})^2$	

10.3.4 Rows and Columns Nested in Replications

Since in Example 10.4 different manufactures are involved, the batches of material and the ovens will be different from one manufacturer (that is, replication) to the next. Hence the factors “batches” and “ovens” are nested within the factor “replications.” An appropriate model then is

$$y_{ijk(l)} = \mu + \alpha_i + \rho_{ij} + \gamma_{ik} + \tau_l + e_{ijk(l)}, \quad (10.30)$$

which leads to the ANOVA given in Table 10.6. In this case it may very well be useful and advisable to amend model (10.30) and include a manufacturer \times treatment interaction term if one wants to investigate whether differences among treatment effects are manufacturer specific due to different production processes employed by the different manufacturers.

In all three cases the hypothesis $\tau_1 = \tau_2 = \dots = \tau_t = 0$ is tested in the usual way by using $F = MS(T)/MS(E)$ with $t - 1$ and ν d.f., where $MS(E)$ and ν are computed differently for the three situations.

10.3.5 Replication \times Treatment Interaction

We have given as a rationale for replicating a basic LSD our desire to increase the number of d.f. for the error SS in order to increase the power of the F -test for treatment effects. But during our discussion in Sections 10.3.2, 10.3.3, 10.3.4 we have already alluded to the fact that such replication may enable us to investigate replication \times treatment interaction. In fact, this may very well be the major reason for replicating the LSD, that is, including another blocking factor, most likely an intrinsic factor, in order to inquire whether the performance of the treatments is the same for the different levels of that intrinsic factor. For this purpose we need to modify the models given above.

Specifically, model (10.28) changes to

$$y_{ijk(l)} = \mu + \alpha_i + \rho_j + \gamma_k + \tau_l + (\alpha\tau)_{il} + e_{ijk(l)}, \quad (10.31)$$

Table 10.6 ANOVA for Model (10.30)

Source	d.f.	SS	E(MS)
Replications	$r - 1$	$t^2 \sum_i (\bar{y}_{i..(\cdot)} - \bar{y}_{...(\cdot)})^2$	
Batches/Reps	$r(t - 1)$	$t \sum_{ij} (\bar{y}_{ij.(\cdot)} - \bar{y}_{i..(\cdot)})^2$	
Ovens/Reps	$r(t - 1)$	$t \sum_{ik} (\bar{y}_{i.k(\cdot)} - \bar{y}_{i..(\cdot)})^2$	
Treatments	$t - 1$	$rt \sum_l (\bar{y}_{... (l)} - \bar{y}_{...(\cdot)})^2$	$\sigma_e^2 + rt \sum_l \tau_l^2 / (t - 1)$
Error	$r(t - 1)(t - 2) + (r - 1)(t - 1)$	Difference	σ_e^2
Total	$rt^2 - 1$	$\sum_{ijk(l)} (y_{ijk(l)} - \bar{y}_{...(\cdot)})^2$	

model 10.29 becomes

$$y_{ijk(l)} = \mu + \alpha_i + \rho_{ij} + \gamma_k + \tau_l + (\alpha\tau)_{il} + e_{ijk(l)}, \quad (10.32)$$

and model 10.30 becomes

$$y_{ijk(l)} = \mu + \alpha_i + \rho_{ij} + \gamma_{ik} + \tau_l + (\alpha\tau)_{il} + e_{ijk(l)}. \quad (10.33)$$

For each case we need to amend the ANOVA tables 10.4, 10.5, and 10.6 by the interaction sum of squares (with A denoting the replication factor)

$$SS(A \times T) = t \sum_{il} (\bar{y}_{i..(l)} - \bar{y}_{i...} - \bar{y}_{... (l)} + \bar{y}_{...})^2$$

with $(r-1)(t-1)$ d.f. As a consequence, the error d.f. will be changed to $r(t-1)(t-2)+2(r-1)(r(t-1)(t-2)+(r-1)(t-1))$, and $r(t-1)(t-2)$, respectively.

To test for interaction we use the F -test

$$F = \frac{MS(A \times T)}{MS(E)}$$

with the appropriate d.f. as given above. To deal with possible interaction we use the same approach as outlined in Section 9.6.8.

10.4 LATIN RECTANGLES

Another method of increasing the error d.f. but still maintaining the Latin square principle is to use a design with rt rows and t columns (or t rows and rt columns as rows and columns are obviously interchangeable). In our example, we may have rt batches

of material available at one time to carry out the experiment, rather than t batches at r different times. An appropriate error-reduction design is then obtained by “inter-mixing” r Latin squares generated by r independent randomizations. Example (ii) in Table 10.1 illustrates such a design. Because of its obvious geometrical appearance and its properties we refer to such a design as a *Latin rectangle*. Obviously each treatment occurs exactly once in each row and r times in each column.

An appropriate linear model for this design is

$$y_{ij(k)} = \mu + \rho_i + \gamma_j + \tau_k + e_{ij(k)}, \quad (10.34)$$

that is, the same as for the LSD except that now $i = 1, 2, \dots, rt$. The ANOVA associated with model (10.34) is given in Table 10.7. Inspection shows that this is very similar to the ANOVA in Table 10.4. Since there are no replications in the sense of Section 10.3 we, of course, cannot obtain a replications \times treatment sum of squares.

10.5 INCOMPLETE LATIN SQUARES

As we have mentioned earlier, one disadvantage of a LSD is that the numbers of rows, columns, and treatments must be the same. This is especially true if the number of treatments is large, since then the heterogeneity of the EUs in the square array may be appreciable as measured by the u_{ij} in (10.3). There exist, however, designs with $r = t$ rows and $c(< t)$ columns which combine the Latin square property of eliminating heterogeneity in two directions with the property of a BIBD of comparing treatments with the same variance. Such designs are referred to as *Youden squares* since they were introduced by Youden (1937) after Yates (1936) considered the special case of $c = t - 1$.

EXAMPLE 10.5: A Youden square for $t = r = 7, c = 3$ is given below (with the treatments numbered $1, 2, \dots, 7$):

Row	Column		
	1	2	3
1	1	2	4
2	2	3	5
3	3	4	6
4	4	5	7
5	5	6	1
6	6	7	2
7	7	1	3

When using this design we would, of course, randomly assign the treatments to the numbers $1, 2, \dots, 7$ and then randomize the rows and columns. It is not difficult to verify that this design with rows as blocks is indeed a special arrangement of the BIBD $(7, 7, 3, 3; 1)$. \square

Table 10.7 ANOVA for Model (10.34)

Source	d.f.	SS	$E(\text{MS})$
Rows	$rt - 1$	$t \sum_i (\bar{y}_{i..(\cdot)} - \bar{y}_{..(\cdot)})^2$	
Columns	$t - 1$	$rt \sum_j (\bar{y}_{.j(\cdot)} - \bar{y}_{..(\cdot)})^2$	
Treatments	$t - 1$	$rt \sum_k (\bar{y}_{..(k)} - \bar{y}_{..(\cdot)})^2$	$\sigma_e^2 + rt \sum_k \tau_k^2 / (t - 1)$
Error	$r(t - 1)(t - 2) + 2(r - 1)(t - 1)$	Difference	σ_e^2
Total	$rt^2 - 1$	$\sum_{i,j(k)} (y_{ij(k)} - \bar{y}_{..(\cdot)})^2$	

More generally, Hartley and Smith (1948) have shown that for all BIBDs with $t = b$ such arrangements exist. A listing of these is given in Cochran and Cox (1957). It should be clear from the description of these designs that the columns are orthogonal to the rows and treatments, but the rows are not orthogonal to the treatments since not every treatment occurs in every row. This implies that, using model (10.16) for the analysis, the estimates of the treatment effects have to be adjusted for row effects, that is, no longer can treatment means be used to estimate treatment effects but LS means must be obtained (see Chapter II.2 and Section II.6.5).

Youden squares can also be used to generate designs with $c > t$ using a method similar to that of constructing extended block designs (see Section 9.8). We simply adjoin to a Latin square (or multiples of Latin squares) a Youden square. Example (iii) in Table 10.1 provides a trivial application of this idea. We may refer to such designs as *extended Latin square designs*.

Finally, the Latin square idea can be modified to include designs with $r = \alpha t$ (α integer), $c < t$ such that the BIBD property holds with rows as blocks and each treatment occurs α times in each column. We call these designs *extended incomplete Latin squares*. A listing of such designs is provided by Cochran and Cox (1957).

10.6 ORTHOGONAL LATIN SQUARES

10.6.1 Græco-Latin Squares

An interesting and sometimes useful generalization of the LSD is obtained by considering elimination of heterogeneity in more than two directions. For elimination of systematic variation in three directions consider the following example.

EXAMPLE 10.6: We want to compare the “usefulness” of four different word processing softwares (A, B, C, D) using four different PCs, four secretaries, and four different texts ($\alpha, \beta, \gamma, \delta$). We want to eliminate differences among PCs, secretaries, and types of text. A suitable arrangement may be as follows:

PC	Secretary			
	1	2	3	4
1	$A\alpha$	$B\gamma$	$C\delta$	$D\beta$
2	$B\beta$	$A\delta$	$D\gamma$	$C\alpha$
3	$C\gamma$	$D\alpha$	$A\beta$	$B\delta$
4	$D\delta$	$C\beta$	$B\alpha$	$A\gamma$

that is, secretary 1 types text α on PC 1 with word processor A , and so forth. \square

This design has interesting combinatorial properties: if we ignore the Greek letters we have a Latin square; if we ignore the Latin letters we also have a Latin square; in addition, each Greek letter occurs exactly once with each Latin letter. We have superimposed two Latin squares on each other with the resulting third property. We refer to such an arrangement as two *orthogonal Latin squares*, or more specifically, as a *Græco-Latin square* (the name is suggested by the use of Greek and Latin letters). Such designs exist for all t except $t = 6$. Græco-Latin squares for $t = 3, 4, 5, 7, 8, 9, 11, 12$ are given by Cochran and Cox (1957), for $t = 10$ by Bose, Shrikhande and Parker (1960), and Fisher and Yates (1957) list complete sets of orthogonal Latin squares for $t = 3, 4, 5, 7, 8, 9$. Pairs from these sets can be used to obtain different Græco-Latin squares (superimposing three orthogonal Latin squares will yield a design suitable to eliminate heterogeneity in four directions, and so on).

A model for analyzing data from a Græco-Latin square design is an obvious extension of (10.16), that is,

$$y_{ij(kl)} = \mu + \rho_i + \gamma_j + \delta_k + \tau_l + e_{ij(kl)} \quad (10.35)$$

where ρ_i are the row effects, γ_j are the column effects, δ_k represent the effects of the blocking factor associated with the Greek letters, and τ_l are the treatment effects ($i, j, k, l = 1, 2, \dots, t$). The fact that out of all possible t^4 combinations (i, j, k, l) only t^2 occur in a Græco-Latin square is indicated in the subscript notation $ij(kl)$ for the observations in model (10.35). As a consequence, after accounting for $t - 1$ d.f. each for rows, columns, Greek letters, and treatments in the ANOVA table (see Table 10.8) only $(t - 1)(t - 3)$ d.f. remain for error. For small t this is usually not sufficient, but matters may be improved through appropriate replication. Yet, the Græco-Latin square suffers from the same (and more) restrictions than the LSD and that may impede its usefulness in practical applications somewhat.

10.6.2 Mutually Orthogonal Latin Squares

The process of superimposing orthogonal Latin squares, and thereby creating error-control designs to eliminate additional sources of variability, can be continued for most values of t . For example, when t is a prime number or a power of a prime number then there exists a $t \times t$ square with each cell containing a letter of $t - 1$ languages, such that

Table 10.8 ANOVA Table for Græco-Latin Square

Source	d.f.	SS	E(MS)
Rows	$t - 1$	$t \sum_i (\bar{y}_{i..} - \bar{y}_{..})^2$	
Columns	$t - 1$	$t \sum_j (\bar{y}_{.j.} - \bar{y}_{..})^2$	
Greek Letters	$t - 1$	$t \sum_k (\bar{y}_{..(k)} - \bar{y}_{..})^2$	
Treatments	$t - 1$	$t \sum_l (\bar{y}_{..(l)} - \bar{y}_{..})^2$	$\sigma_e^2 + t \sum_l \tau_l^2 / (t - 1)$
Error	$(t - 1)(t - 3)$	Difference	σ_e^2
Total	$t^2 - 1$	$\sum_{ijkl} (y_{ijkl} - \bar{y}_{..})^2$	

the letters of any two languages form a square with the Græco-Latin square property (Bose, 1938; Stevens, 1938). Consider the following example.

EXAMPLE 10.7: For $t = 4$ with the $t - 1 = 3$ languages being Latin letters, Greek letters, and numerals, an arrangement of three orthogonalized squares is given by

$$\begin{array}{cccc} A\alpha 1 & B\beta 2 & C\gamma 3 & D\delta 4 \\ B\gamma 4 & A\delta 3 & D\alpha 2 & C\beta 1 \\ C\delta 2 & D\gamma 1 & A\beta 4 & B\alpha 3 \\ D\beta 3 & C\alpha 4 & B\delta 1 & A\gamma 2 \end{array}$$

□

Such squares are referred to as *completely orthogonalized squares*. In general the, say, $k \leq t - 1$ superimposed Latin squares with the property described above are called *mutually orthogonal Latin squares* (MOLS). As error-control designs they can be used to eliminate then $k + 2$ sources of variation.

10.7 CHANGE-OVER DESIGNS

The structure of a Latin square forms the basis for a variety of error-control designs. In this section we shall discuss briefly such a situation where individuals (subjects) are used as one blocking factor and time period as the other blocking factor. These designs have been used extensively in different kinds of experimental settings, but mainly in the pharmaceutical industry during the testing of new drugs, in animal science for feeding experiments, and in psychological studies. The basic idea is that each individual receives (sequentially) all or some of the treatments, one at any given time period, and that for different individuals the order of the application of the treatments is being changed. And even though the designs for different applications have the same features, they are often referred to by different names, such as *cross-over designs*, *change-over designs*, *carry-over designs*, *switch-over designs*, *counter-balanced designs*, and

sometimes more generally and generically by *repeated measurement designs* (see also Chapter 14). Also, because there often exists considerable variability among subjects the fact that each subject is exposed to every treatment is described as “each subject being its own control.”

10.7.1 Two-Treatment Change-Over Design

In its simplest form a change-over design consists of $n = 2r$ subjects to compare two treatments, A and B say, over two time periods. In this situation r subjects receive the treatments in the sequence $A - B$, that is, treatment A in period 1 and treatment B in period 2. The remaining r subjects receive the treatments in sequence $B - A$, that is, in reverse order. With periods as “rows” and subjects as “columns” this design has obviously the form of a Latin rectangle (see Section 10.4), intermixing $r \times 2$ Latin squares.

The usefulness of this design depends on whether:

- (i) we assume that because of a suitable “wash-out period” between periods 1 and 2 the observations in period 2 are affected only by the treatment applied in period 2, that is, the treatment applied in period 1 has no effect on the outcome, or
- (ii) there is a *carry-over effect* from period 1 to period 2, that is, the treatment applied in period 1 has an effect on the observation in period 2, in addition to the effect of the treatment applied in period 2.

An appropriate model for situation (i) is

$$y_{ij(k)} = \mu + p_i + s_j + \tau_k + e_{ij(k)}, \quad (10.36)$$

where p_i is the effect of the i th period and s_j is the effect of the j th subject with $i = 1, 2; j = 1, 2, \dots, 2r$. Model (10.36) is, of course, the same as model (10.34).

For situation (ii) model (10.36) has to be amended for observations in period 2 to account for the carry-over or *residual effects*. Suppose that subject 1 receives the treatments in the sequence $A - B$, and subject 2 receives the treatments in the sequence $B - A$. Then the observations $y_{11A}, y_{21B}, y_{12B}, y_{22A}$ can be modelled as follows:

$$\begin{aligned} y_{11A} &= \mu + p_1 + s_1 + \tau_A & + e_{11A} \\ y_{21B} &= \mu + p_2 + s_1 + \tau_B + \gamma_A + e_{21B} \end{aligned} \quad (10.37)$$

$$\begin{aligned} y_{12B} &= \mu + p_1 + s_2 + \tau_B & + e_{12B} \\ y_{22A} &= \mu + p_2 + s_2 + \tau_A + \gamma_B + e_{22A}, \end{aligned} \quad (10.38)$$

where γ_A and γ_B represent the residual effects for treatments A and B , respectively. To make the distinction clear, τ_A and τ_B are also referred to as the *direct effects* of treatments A and B , respectively. More generally we may write models 10.37 and 10.38 as

$$y_{ij(kl)} = \mu + p_i + s_j + \tau_k + \gamma_{(l)} + e_{ij(kl)}, \quad (10.39)$$

where τ_k refers to the direct effect of the treatment assigned to subject j in period i , and $\gamma_{(l)}$ represents the residual or carry-over effect of the treatment assigned to subject

j in period $i - 1$, with $\gamma_{(l)} = 0$ for $i = 1$, that is, there is no carry-over effect for observations in period 1.

Returning to our basic design which is essentially of the form

Period	Sequence	
	1	2
1	A	B
2	B	A

with the two sequences repeated r times. For purposes of parameter estimation it is sufficient to simply consider the case $r = 1$. We thus have four observations, that is, three degrees of freedom. We know already that under (i) above and model (10.36) we attribute one d.f. each to periods, subjects (sequences), and treatments (see Table 10.8). Repeating the sequences merely provides error d.f. (see Table 10.8).

Considering now situation (ii) above and model (10.39) we recognize immediately that the suggested design does not provide enough d.f. to estimate unbiasedly contrasts for all four sets of parameters. Thus, another type of design is required. One such design with two periods was proposed by Balaam (1968), although he originally considered situation (i) above with the possibility of including period \times treatment interaction. His design consists of r repeats of the following sequences:

Period	Sequence			
	1	2	3	4
1	A	B	A	B
2	B	A	A	B

Using model (10.39) the expected values of the eight means from this design can be written as

$$\mu_{11A} = \mu + p_1 + s_1 + \tau_A$$

$$\mu_{21B} = \mu + p_2 + s_1 + \tau_B + \gamma_A$$

$$\mu_{12B} = \mu + p_1 + s_2 + \tau_B$$

$$\mu_{22A} = \mu + p_2 + s_2 + \tau_A + \gamma_B$$

$$\mu_{13A} = \mu + p_1 + s_3 + \tau_A$$

$$\mu_{23A} = \mu + p_2 + s_3 + \tau_A + \gamma_A$$

$$\mu_{14B} = \mu + p_1 + s_4 + \tau_B$$

$$\mu_{24B} = \mu + p_2 + s_4 + \tau_B + \gamma_B$$

Contrasts among these μ_{ijk} can then be used to obtain functions of $\tau_A - \tau_B$ and $\gamma_A - \gamma_B$

only. Specifically, if we write

$$\mu_{11A} - \mu_{21B} \equiv \textcircled{1}$$

$$\mu_{12B} - \mu_{22A} \equiv \textcircled{2}$$

$$\mu_{13A} - \mu_{23A} \equiv \textcircled{3}$$

$$\mu_{14B} - \mu_{24B} \equiv \textcircled{4}$$

we then find that

$$\textcircled{4} - \textcircled{3} = \gamma_A - \gamma_B$$

and

$$\textcircled{1} - \textcircled{2} - \textcircled{4} + \textcircled{3} = 2(\tau_A - \tau_B).$$

Other possible designs may use three or more periods rather than two. Examples of three-period designs are

Period	Sequence			Period	Sequence	
	1	2			1	2
1	A	B	or	1	A	B
2	B	A		2	B	A
3	A	B		3	B	A

repeating each sequence r times. For still other designs and more details see Jones and Kenward (2003) or Section II.19.5.8.

For analysis purposes we may rewrite model 10.39 in matrix form as

$$\mathbf{y} = \mu\mathbf{J} + \mathbf{X}_p\mathbf{p} + \mathbf{X}_s\mathbf{s} + \mathbf{X}_\tau\boldsymbol{\tau} + \mathbf{X}_\gamma\boldsymbol{\gamma} + \mathbf{e} \quad (10.40)$$

using obvious notation and apply the general methods discussed in Chapter 4. More precisely, (10.40) is a 5-part linear model and we can write down the full set of NEs or obtain the reduced NE for $\boldsymbol{\tau}$ and $\boldsymbol{\gamma}$ and solve for the desired effects. Furthermore, we can obtain $SS(\mathbf{X}_\tau|\mathbf{J}, \mathbf{X}_p, \mathbf{X}_s, \mathbf{X}_\gamma)$ and $SS(\mathbf{X}_\gamma|\mathbf{J}, \mathbf{X}_p, \mathbf{X}_s, \mathbf{X}_\tau)$ to test hypotheses concerning the direct and residual treatment effects, respectively.

Lucas (1957) and Cochran and Cox (1957) have pointed out that, because of the nonorthogonality of change-over designs, residual effects are estimated less precisely than direct effects. To alleviate this problem and to achieve orthogonality between direct and residual effects Lucas (1957) suggested to add an extra-period, referred to as preperiod, to the basic design, that is, change, for example

Period	Sequence			Period	Sequence	
	1	2			1	2
1	A	B	to	0	A	B
2	B	A		1	A	B
3	A	B		2	B	A
				3	A	B

without taking observations in period 0, the preperiod. In the augmented design each combination of direct and residual effects occur exactly once (or r times for the entire design), namely $(\tau_A, \gamma_A), (\tau_A, \gamma_B), (\tau_B, \gamma_A), (\tau_B, \gamma_B)$, implying orthogonality. This in turn simplifies the analysis as sketched above.

10.7.2 Change-Over Designs for More than Two Treatments

For the general case of $t(> 2)$ treatments in a change-over design Williams (1949, 1950) showed how Latin squares can be used to obtain what he called *balanced residual effects designs*. The basic properties of these designs are that each treatment occurs the same number of times, λ_1 say, and each treatment is preceded by every other treatment the same number of times, say λ_2 (this is actually a special case of the more general definition of a balanced repeated measurement design given by Hedayat and Afsarinejad, 1975). These designs consist of one cyclic Latin square if t is even and of two cyclic Latin squares if t is odd.

EXAMPLE 10.8: For $t = 6$ the design with t periods and t subjects is as follows (designating the treatments by numbers rather than Latin letters):

Period	Subject					
	1	2	3	4	5	6
1	1	2	3	4	5	6
2	6	1	2	3	4	5
3	2	3	4	5	6	1
4	5	6	1	2	3	4
5	3	4	5	6	1	2
6	4	5	6	1	2	3

□

This example illustrates the general method of constructing these designs. For subject 1 the treatments $1, 2, \dots, t/2$ occur in the periods $1, 3, \dots, t - 1$, respectively, and the treatments $t/2 + 1, t/2 + 2, \dots, t$ occur in the periods $t, t - 2, \dots, 2$, respectively. The assignments for the remaining subjects are obtained by simply adding $1, 2, \dots, t -$

1 to the treatments for subject 1 (with reduction modulo t) for subjects $2, 3, \dots, t$, respectively.

For t odd we use two cyclic Latin squares for $2t$ subjects and t periods.

EXAMPLE 10.9: For $t = 5$ the design is as follows:

Period	Subject									
	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	2	3	4	5	1
2	5	1	2	3	4	3	4	5	1	2
3	2	3	4	5	1	1	2	3	4	5
4	4	5	1	2	3	4	5	1	2	3
5	3	4	5	1	2	5	1	2	3	4

□

Here treatments $1, 2, \dots, (t+1)/2$ occur in periods $1, 3, \dots, t$, respectively, and treatments $(t+1)/2 + 1, (t+1)/2 + 2, \dots, t$ in periods $t-1, t-3, \dots, 2$, respectively, for subject 1. The assignments for subjects $2, 3, \dots, t$ are obtained through a cyclic development of the arrangement for subject 1 as described previously. The arrangement for subject $t+1$ is the mirror image, that is, reverse order, of the arrangement for subject t , and the assignment for the remaining subjects is again obtained through cyclic development. For a more general discussion see Section II.19.5.1.

In the behavioral science literature these designs are often referred to as *completely counterbalanced Latin squares* (Wagenaar, 1969) (see also Section 13.4). This does not mean, however, that for these designs the direct and residual effects are orthogonal to each other. Orthogonality can be achieved, as discussed earlier, by adding a preperiod with the same treatment arrangement as in period 1, so that every treatment is also preceded λ_2 times by itself. With or without the preperiod an appropriate model is of the form of (10.39).

10.7.3 Some Variations and Extensions

There exist, obviously, many variations and extensions of these designs. For example we may have p , the number of periods, less than t , the number of treatments. This may occur when the number of treatments is large and, because of fatigue, not each participant can be assigned each treatment, or assigning each treatment to each participant may simply take too much time, in particular if a sufficient number of participants is available for the experiment. In this situation the subjects may represent some form of incomplete block, and the basic building block may be incomplete Latin squares (Patterson, 1950, 1951, 1952). Afsarinejad (1990) has extended the algorithm for the Williams designs (Examples 10.8 and 10.9) to construct balanced designs for the case $p < t$ (see also Section II.19.5.2). We shall not provide details here, but give the following example.

EXAMPLE 10.10: For $t = 5$ and $p = 3$ the following design is a balanced design with 10 subjects in which each treatment is preceded and followed by every other treatment exactly once:

Period	Subject									
	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	3	4	5	1	2
2	5	1	2	3	4	5	1	2	3	4
3	3	4	5	1	2	1	2	3	4	5

□

We may not have enough subjects for a balanced design as described above. This has led to the development of partially balanced designs (Patterson and Lucas, 1962) using PBIB designs for purposes of construction (see also Blaisdell and Raghavarao, 1980) (see Sections II.19.5.2 and 19.5.3).

An alternative design, proposed by Balaam (1968), uses only two periods but t^2 subjects for t treatments. The basic idea is to assign all $t(t-1)$ ordered pairs of treatments to $t(t-1)$ subjects for periods 1 and 2 and use t subjects receiving the same treatment in both periods, as illustrated in the following example.

EXAMPLE 10.11: For $t = 3$ the Balaam (1968) design is given by

Period	Subjects								
	1	2	3	4	5	6	7	8	9
1	A	A	B	B	C	C	A	B	C
2	B	C	C	A	A	B	A	B	C

□

An extension of the designs discussed so far leads to designs balanced for second order residual effects (Williams, 1949, 1950). In this case we consider carry-over effects not just from the immediately preceding treatment but also from the treatment applied two periods prior to the present application. For such a situation model (10.40) needs to be amended by an additional term representing the second order residual effect. Obviously the construction of such designs becomes more complicated. Williams (1949, 1950) showed, for example, how mutually orthogonal Latin squares can be used to achieve the goal.

EXAMPLE 10.12: For $t = 4$, $p = 4$ and $s = 12$ subjects Williams gives the following design:

Period	Subjects				Subjects				Subjects			
	1	2	3	4	5	6	7	8	9	10	11	12
1	4	3	2	1	4	3	2	1	4	3	2	1
2	1	2	3	4	3	4	1	2	2	1	4	3
3	2	1	4	3	1	2	3	4	3	4	1	2
4	3	4	1	2	2	1	4	3	1	2	3	4

The distinctive feature of this design is that each treatment is preceded exactly once by each ordered pair of treatments. For more details see Section II.19.8.5. \square

10.8 EXAMPLES USING SAS®

EXAMPLE 10.13: Consider an agricultural experiment with $t = 5$ treatments in a Latin square design layout due to fertility gradients in two directions. The design and the data are given in Table 10.9a.

Treatment 1 represents a control and the main objective is to compare treatments 2, 3, 4, 5 versus treatment 1. The input statements using SAS PROC GLM are given in Table 10.9a. In addition to considering the comparison of treatment 1 versus the average of the remaining four treatments, we want to perform Dunnett's procedure (see Section 7.5.7).

The results are given in Table 10.9b:

- (i) The overall treatment differences are significant at $P = .0523$.
- (ii) Dunnett's procedure shows treatments 2 and 5 are clearly significantly different from treatment 1 ($P = .0459$ and $P = .0218$, respectively), whereas treatment 3 is marginally significantly different from treatment 1 ($P = .1138$)
- (iii) Having specified $\alpha = .10$, 90% simultaneous confidence intervals for $\tau_i - \tau_1$ ($i = 2, 3, 4, 5$) are provided.
- (iv) The estimate for $(\tau_2 + \tau_3 + \tau_4 + \tau_5)/4 - \tau_1$ is 20.15, indicating that on the average the new treatments provide a higher yield than the control. \square

EXAMPLE 10.14: A multi-farm trial was performed to evaluate the effectiveness of different doses, low (L), medium (M), high (H), of a food additive on growth of cattle. In addition to the three doses a control (C) was included in the experiment. The investigator decided to block on farms and weight classes, leading to a 4×4 Latin square design. Two breeds were included in the experiments, using four farms for each breed.

Table 10.9 Latin Square Design

a) Input statements:

```
data weight;
data LatinSq;
input row column trt y @@;
datalines;
1 1 1 94 1 2 3 100 1 3 4 98 1 4 2 101 1 5 5 112
2 1 3 103 2 2 2 111 2 3 1 51 2 4 5 110 2 5 4 90
3 1 4 114 3 2 1 75 3 3 5 94 3 4 3 85 3 5 2 107
4 1 5 100 4 2 4 74 4 3 2 70 4 4 1 93 4 5 3 106
5 1 2 106 5 2 5 95 5 3 3 81 5 4 4 90 5 5 1 73
;
run;

proc glm data=LatinSq;
class row column trt;
model y=row column trt;
lsmeans trt/stderr pdiff cl adjust=Dunnett alpha=.10;
estimate '1 vs (2+3+4+5)' trt -4 1 1 1 1 /divisor=4;
title1 'LATIN SQUARE DESIGN';
title2 'ANALYSIS OF VARIANCE';
title3 'W/POST-HOC ANALYSIS';
run;
```

b.) Output:

LATIN SQUARE DESIGN ANALYSIS OF VARIANCE W/POST-HOC ANALYSIS					
The GLM Procedure					
Class Level Information					
Class	Levels	Values			
row	5	1	2	3	4 5
column	5	1	2	3	4 5
trt	5	1	2	3	4 5
Number of Observations Read					25
Number of Observations Used					25
Dependent Variable: y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	4094.720000	341.226667	2.34	0.0774
Error	12	1748.720000	145.726667		
Corrected Total	24	5843.440000			
	R-Square	Coeff Var	Root MSE	y Mean	
	0.700738	12.93584	12.07173	93.32000	

Table 10.9 (Continued)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
row	4	514.240000	128.560000	0.88	0.5033
column	4	1711.440000	427.860000	2.94	0.0661
trt	4	1869.040000	467.260000	3.21	0.0523

Source	DF	Type III SS	Mean Square	F Value	Pr > F
row	4	514.240000	128.560000	0.88	0.5033
column	4	1711.440000	427.860000	2.94	0.0661
trt	4	1869.040000	467.260000	3.21	0.0523

Least Squares Means
Adjustment for Multiple Comparisons: Dunnett

trt	y LSMEAN	Standard Error	H0:LSMEAN=0 Pr > t	H0:LSMean=Control Pr > t
1	77.200000	5.398642	<.0001	
2	99.000000	5.398642	<.0001	0.0459
3	95.000000	5.398642	<.0001	0.1138
4	93.200000	5.398642	<.0001	0.1680
5	102.200000	5.398642	<.0001	0.0218

trt	y LSMEAN	90% Confidence Limits	
1	77.200000	67.578068	86.821932
2	99.000000	89.378068	108.621932
3	95.000000	85.378068	104.621932
4	93.200000	83.578068	102.821932
5	102.200000	92.578068	111.821932

Least Squares Means for Effect trt

i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
2	1	21.800000	3.415303	40.184697
3	1	17.800000	-0.584697	36.184697
4	1	16.000000	-2.384697	34.384697
5	1	25.000000	6.615303	43.384697

Dependent Variable: y

Parameter	Estimate	Standard Error	t Value	Pr > t
1 vs (2+3+4+5)	20.150000	6.03586503	3.34	0.0059

Thus, we have a replicated LS design with farms nested in breeds and weight classes crossed with breeds.

The design and the data are given in Table 10.10a and the results of the analysis, using SAS PROC GLM, are given in Table 10.10b:

- (i) Differences among dosages and interaction between breed and dosage were highly significant ($P < .0001$).
- (ii) The LS means for breed*dosage indicate that for both breeds growth remains nearly the same for *C*, *L*, *M*, but increases for *H*, with a substantially higher increase for breed 2, leading to the significant interaction. The interaction is, however, co-directional. Hence it is meaningful to assert that on average, dosage *H* leads to higher growth by about 10 kg, 153 vs 163.
- (iii) The slice operator performs separate analyses (but using the same error term, $MS(Error) = 1.93$) and concludes that for both breeds the differences among dosages are highly significant, due, of course, to the performance of *H*. \square

EXAMPLE 10.15: Consider the following crossover design with $t = 3$ treatments and $p = 3$ periods. There are six possible sequences of assigning the treatments over the three periods used in the experiment. Each sequence is replicated twice. The plan and the data are given in Table 10.11a.

We use SAS PROC GLM to analyze the data. The input statements are given in Table 10.11a. Since for the observations in period 1 there are carry-over effects we put a “0” in the column for carry-over effects. Since this will result in “non-estimable” LS means for treatments we follow Ratkowsky, Evans and Alldredge (1993) to insert the statement “if carry = ‘0’ then carry = ‘3’ ” to alleviate this problem (we shall comment on this below).

The analysis of the data is given in Table 10.11b:

- (i) The general form of estimable function shows that differences between treatment and carry-over effects are estimable.
- (ii) There are highly significant differences among the treatments ($P < .0001$).
- (iii) Differences among carry-over effects are not significant ($P = .41$).
- (iv) The coefficients for trt and carry LS means show how those LS means are obtained from the solution vector (not shown here).
- (v) The treatment LS means together with their standard errors:

56.71	±	1.05
52.69	±	1.05
47.14	±	.85

Table 10.10 Replicated Latin Square Design

a) Input statements:

```
data repLS
input breed farm wclass dosage $ response @@;
datalines;
1 1 1 H 142 1 1 2 L 144 1 1 3 C 148 1 1 4 M 150
1 2 1 M 138 1 2 2 C 143 1 2 3 L 145 1 2 4 H 154
1 3 1 C 138 1 3 2 M 144 1 3 3 H 153 1 3 4 L 149
1 4 1 L 139 1 4 2 H 149 1 4 3 M 144 1 4 4 C 150
2 1 1 L 158 2 1 2 C 161 2 1 3 M 165 2 1 4 H 180
2 2 1 C 155 2 2 2 L 160 2 2 3 H 178 2 2 4 M 167
2 3 1 M 158 2 3 2 H 175 2 3 3 L 163 2 3 4 C 167
2 4 1 H 174 2 4 2 M 161 2 4 3 C 164 2 4 4 L 168
;
run;

proc glm data=repLS;
class breed farm wclass dosage;
model response= breed farm(breed) wclass dosage breed*dosage;
lsmeans dosage/stderr;
lsmeans breed*dosage/stderr slice=breed;
title1 'REPLICATED LATIN SQUARE DESIGN';
title2 '(r=2, t=4, Rows Nested in Replications)';
title3 'ANALYSIS OF VARIANCE';
run;
```

b.)Output:

REPLICATED LATIN SQUARE DESIGN (r=2, t=4, Rows Nested in Replications) ANALYSIS OF VARIANCE					
The GLM Procedure					
Class Level Information					
Class	Levels	Values			
breed	2	1 2			
farm	4	1 2 3 4			
wclass	4	1 2 3 4			
dosage	4	C H L M			
Number of Observations Read					32
Number of Observations Used					32
Dependent Variable: response					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	4470.250000	279.390625	140.87	<.0001
Error	15	29.750000	1.983333		
Corrected Total	31	4500.000000			

Table 10.10 (Continued)

R-Square					
Coeff Var					
Root MSE					
response Mean					
0.993389					
0.904211					
1.408309					
155.7500					
Source					
DF					
Type I SS					
Mean Square					
F Value					
Pr > F					
breed					
1					
3280.500000					
3280.500000					
1654.03					
<.0001					
farm(breed)					
6					
9.000000					
1.500000					
0.76					
0.6146					
wclass					
3					
466.750000					
155.583333					
78.45					
<.0001					
dosage					
3					
580.250000					
193.416667					
97.52					
<.0001					
breed*dosage					
3					
133.750000					
44.583333					
22.48					
<.0001					
Source					
DF					
Type III SS					
Mean Square					
F Value					
Pr > F					
breed					
1					
3280.500000					
3280.500000					
1654.03					
<.0001					
farm(breed)					
6					
9.000000					
1.500000					
0.76					
0.6146					
wclass					
3					
466.750000					
155.583333					
78.45					
<.0001					
dosage					
3					
580.250000					
193.416667					
97.52					
<.0001					
breed*dosage					
3					
133.750000					
44.583333					
22.48					
<.0001					
Least Squares Means					
dosage					
response					
LSMEAN					
Standard					
Error					
Pr > t					
C					
153.250000					
0.497912					
<.0001					
H					
163.125000					
0.497912					
<.0001					
L					
153.250000					
0.497912					
<.0001					
M					
153.375000					
0.497912					
<.0001					
breed					
dosage					
response					
LSMEAN					
Standard					
Error					
Pr > t					
1					
C					
144.750000					
0.704154					
<.0001					
1					
H					
149.500000					
0.704154					
<.0001					
1					
L					
144.250000					
0.704154					
<.0001					
1					
M					
144.000000					
0.704154					
<.0001					
2					
C					
161.750000					
0.704154					
<.0001					
2					
H					
176.750000					
0.704154					
<.0001					
2					
L					
162.250000					
0.704154					
<.0001					
2					
M					
162.750000					
0.704154					
<.0001					
breed*dosage Effect Sliced by breed for response					
breed					
DF					
Sum of					
Squares					
Mean Square					
F Value					
Pr > F					
1					
3					
81.250000					
27.083333					
13.66					
0.0001					
2					
3					
632.750000					
210.916667					
106.34					
<.0001					

Table 10.11 Crossover Design

a) Input statements:

```
data cross;
input period sequence steer trt carry y @@;
if carry='0' then carry= '3';
datalines;
1 1 1 1 0 50 2 1 1 2 1 61 3 1 1 3 2 53
1 1 2 1 0 55 2 1 2 2 1 63 3 1 2 3 2 57
1 2 3 2 0 44 2 2 3 3 2 42 3 2 3 1 3 57
1 2 4 2 0 51 2 2 4 3 2 46 3 2 4 1 3 59
1 3 5 3 0 35 2 3 5 1 3 55 3 3 5 2 1 47
1 3 6 3 0 41 2 3 6 1 3 56 3 3 6 2 1 50
1 4 7 1 0 54 2 4 7 3 1 48 3 4 7 2 3 51
1 4 8 1 0 58 2 4 8 3 1 51 3 4 8 2 3 54
1 5 9 2 0 50 2 5 9 1 2 57 3 5 9 3 1 51
1 5 10 2 0 55 2 5 10 1 2 59 3 5 10 3 1 55
1 6 11 3 0 41 2 6 11 2 3 56 3 6 11 1 2 58
1 6 12 3 0 46 2 6 12 2 3 58 3 6 12 1 2 61
;
run;

proc glm data=cross;
class period sequence steer trt carry;
model y=period sequence steer(sequence) trt carry/e;
lsmeans trt carry/stderr e;
estimate '1-2' trt 1 -1 0;
estimate '1-3' trt 1 0 -1;
estimate '2-3' trt 0 1 -1;
title1 'CROSSOVER DESIGN';
title2 'USING COUNTERBALANCED LATIN SQUARES';
run;
```

b.) Output:

CROSSOVER DESIGN		
USING COUNTERBALANCED LATIN SQUARES		
The GLM Procedure		
Class Level Information		
Class	Levels	Values
period	3	1 2 3
sequence	6	1 2 3 4 5 6
steer	12	1 2 3 4 5 6 7 8 9 10 11 12
trt	3	1 2 3
carry	3	1 2 3
Number of Observations Read		36

Table 10.11 (Continued)

General Form of Estimable Functions					
Effect			Coefficients		
Intercept			L1		
period	1		L2		
period	2		L3		
period	3		L1-L2-L3		
sequence	1		L5		
sequence	2		L6		
sequence	3		L7		
sequence	4		L8		
sequence	5		L9		
sequence	6		L1-L5-L6-L7-L8-L9		
steer(sequence)	1	1	L11		
steer(sequence)	2	1	L5-L11		
steer(sequence)	3	2	L13		
steer(sequence)	4	2	L6-L13		
steer(sequence)	5	3	L15		
steer(sequence)	6	3	L7-L15		
steer(sequence)	7	4	L17		
steer(sequence)	8	4	L8-L17		
steer(sequence)	9	5	L19		
steer(sequence)	10	5	L9-L19		
steer(sequence)	11	6	L21		
steer(sequence)	12	6	L1-L5-L6-L7-L8-L9-L21		
trt	1		L23		
trt	2		L24		
trt	3		L1-L23-L24		
carry	1		L26		
carry	2		L27		
carry	3		L1-L26-L27		

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	1302.513889	76.618464	8.74	<.0001
Error	18	157.791667	8.766204		
Corrected Total	35	1460.305556			

R-Square	Coeff Var	Root MSE	y Mean
0.891946	5.654535	2.960778	52.36111

Table 10.11 (Continued)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
period	2	292.0555556	146.0277778	16.66	<.0001
sequence	5	326.4722222	65.2944444	7.45	0.0006
steer(sequence)	6	118.5000000	19.7500000	2.25	0.0849
trt	2	549.0555556	274.5277778	31.32	<.0001
carry	2	16.4305556	8.2152778	0.94	0.4100

Source	DF	Type III SS	Mean Square	F Value	Pr > F
period	2	172.3072917	86.1536458	9.83	0.0013
sequence	5	318.6916667	63.7383333	7.27	0.0007
steer(sequence)	6	118.5000000	19.7500000	2.25	0.0849
trt	2	440.6083333	220.3041667	25.13	<.0001
carry	2	16.4305556	8.2152778	0.94	0.4100

Coefficients for trt Least Square Means

Effect		trt Level		
		1	2	3
Intercept		1	1	1
period	1	0.33333333	0.33333333	0.33333333
period	2	0.33333333	0.33333333	0.33333333
period	3	0.33333333	0.33333333	0.33333333
sequence	1	0.16666667	0.16666667	0.16666667
sequence	2	0.16666667	0.16666667	0.16666667
sequence	3	0.16666667	0.16666667	0.16666667
sequence	4	0.16666667	0.16666667	0.16666667
sequence	5	0.16666667	0.16666667	0.16666667
sequence	6	0.16666667	0.16666667	0.16666667
steer(sequence)	1 1	0.08333333	0.08333333	0.08333333
steer(sequence)	2 1	0.08333333	0.08333333	0.08333333
steer(sequence)	3 2	0.08333333	0.08333333	0.08333333
steer(sequence)	4 2	0.08333333	0.08333333	0.08333333
steer(sequence)	5 3	0.08333333	0.08333333	0.08333333
steer(sequence)	6 3	0.08333333	0.08333333	0.08333333
steer(sequence)	7 4	0.08333333	0.08333333	0.08333333
steer(sequence)	8 4	0.08333333	0.08333333	0.08333333
steer(sequence)	9 5	0.08333333	0.08333333	0.08333333
steer(sequence)	10 5	0.08333333	0.08333333	0.08333333
steer(sequence)	11 6	0.08333333	0.08333333	0.08333333
steer(sequence)	12 6	0.08333333	0.08333333	0.08333333
trt	1	1	0	0
trt	2	0	1	0
trt	3	0	0	1
carry	1	0.33333333	0.33333333	0.33333333
carry	2	0.33333333	0.33333333	0.33333333
carry	3	0.33333333	0.33333333	0.33333333

Table 10.11 (Continued)

	trt	y LSMEAN	Standard Error	Pr > t
	1	56.7083333	1.0467929	<.0001
	2	52.6875000	1.0467929	<.0001
	3	47.1666667	0.8547029	<.0001

Coefficients for carry Least Square Means

Effect		carry Level		
		1	2	3
Intercept		1	1	1
period	1	0.33333333	0.33333333	0.33333333
period	2	0.33333333	0.33333333	0.33333333
period	3	0.33333333	0.33333333	0.33333333
sequence	1	0.16666667	0.16666667	0.16666667
sequence	2	0.16666667	0.16666667	0.16666667
sequence	3	0.16666667	0.16666667	0.16666667
sequence	4	0.16666667	0.16666667	0.16666667
sequence	5	0.16666667	0.16666667	0.16666667
sequence	6	0.16666667	0.16666667	0.16666667
steer(sequence)	1 1	0.08333333	0.08333333	0.08333333
steer(sequence)	2 1	0.08333333	0.08333333	0.08333333
steer(sequence)	3 2	0.08333333	0.08333333	0.08333333
steer(sequence)	4 2	0.08333333	0.08333333	0.08333333
steer(sequence)	5 3	0.08333333	0.08333333	0.08333333
steer(sequence)	6 3	0.08333333	0.08333333	0.08333333
steer(sequence)	7 4	0.08333333	0.08333333	0.08333333
steer(sequence)	8 4	0.08333333	0.08333333	0.08333333
steer(sequence)	9 5	0.08333333	0.08333333	0.08333333
steer(sequence)	10 5	0.08333333	0.08333333	0.08333333
steer(sequence)	11 6	0.08333333	0.08333333	0.08333333
steer(sequence)	12 6	0.08333333	0.08333333	0.08333333
trt	1	0.33333333	0.33333333	0.33333333
trt	2	0.33333333	0.33333333	0.33333333
trt	3	0.33333333	0.33333333	0.33333333
carry	1	1	0	0
carry	2	0	1	0
carry	3	0	0	1

carry	y LSMEAN	Standard Error	Pr > t
1	53.0833333	1.3514039	<.0001
2	50.7708333	1.3514039	<.0001
3	52.7083333	0.8547029	<.0001

Dependent Variable: y

Parameter	Estimate	Standard Error	t Value	Pr > t
1-2	4.02083333	1.35140388	2.98	0.0081
1-3	9.54166667	1.35140388	7.06	<.0001
2-3	5.52083333	1.35140388	4.09	0.000

intuitively confirm the differences among the treatment effects. The values for the LS means are, however, not unique because of the arbitrary choice to replace $\text{carry} = 0$ by $\text{carry} = 3$ (see above). This also affects the standard errors (that is, 1.05 vs. .85). In spite of this, differences between LS means are unique.

- (vi) Differences between treatment are all significantly different from zero. And the estimates of those differences have the same standard error. \square

10.9 EXERCISES

10.1 A marketing expert for a publishing house wants to measure reader preference for three different covers of the same paperback novel. She has chosen 10 cities and 3 newsstands in each city which are going to sell the novel. She wants to use one of two experimental setups described below.

- (a) In each city each cover is assigned randomly to one of the 3 newsstands. The number of books sold during a three-week period following the assignment is used to compare the effect of the covers on sale of the novel.
- (b) In each city each of the 3 newsstands will sell the book using each cover for one week (that is, the trial extends over 3 weeks) in such a way that during a given week the 3 newsstands in a city will display the book with a different cover. The same 3-week period will be used in all cities. Sales figures for each week will be used for the analysis.

For each of the two scenarios described above:

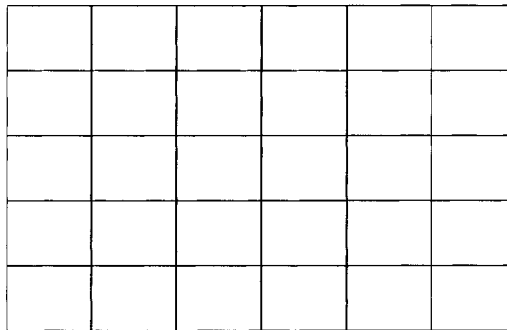
- (i) Give the name of the experimental design used.
 - (ii) Identify the experimental units.
 - (iii) Give the model for each of the designs and the ANOVA table, including sources of variation and d.f.
 - (iv) Indicate how you would test whether the covers had the same effect on sales.
 - (v) Which of the two designs would you prefer in this situation and why?
- 10.2** A study is planned to investigate (a) whether four gasoline additives differ with respect to the reduction in oxides of nitrogen and (b) if such differences exist whether they depend on the makes of the cars used in the study.

The investigator has selected three makes (models) of cars, Ford, Honda, and Porsche. For each model he has four cars available, and he uses four different drivers. He believes that for each model systematic differences are likely to occur in the cars' performance. Also, even though the drivers may do their best to drive the car in a manner required by the test, systematic differences can occur from driver to driver.

In planning the study the investigator would like to have an experimental design that eliminates the car-to-car variation and the driver-to-driver variation. He wants to use the same four drivers for the whole experiment.

- (i) Give the name for an appropriate experimental design.
- (ii) Write out the actual assignment of the additives to car-driver combinations for the three models.
- (iii) Give a linear model and outline the ANOVA table for this experiment, giving sources of variation and d.f.
- (iv) Give the SAS statements (classes, model) for the model in (iii).
- (v) Indicate how you would use the ANOVA to investigate the questions (a) and (b) raised above.

- 10.3** Suppose a poultry scientist comes to you to help him set up an experiment. He wants to compare the effects of 3 different diets (treatments) on eggshell properties. He has available 6 strains of chickens. Each chicken included in the experiment will be housed in a separate pen during the duration of the trial. He has 30 pens available which are arranged in stacks of 5 side-by-side (see diagram below).



For each chicken, measurements are taken on 5 randomly selected eggs.

- (i) What kind of experimental design would you use? Give its parameters (that is, t , b , etc.)?
- (ii) Give a suitable arrangement of the diets (A , B , C) to the chickens.
- (iii) For the design proposed in (ii) give an appropriate linear model and outline the ANOVA, giving sources of variation and d.f.
- (iv) Upon further questioning you find out that the height of the pen in the stack may have an effect on the outcome of the experiment (because of differences in the temperature). Would you change the arrangement of the treatments given in (ii)? If your answer is “no”, give reasons for it; if your answer is “yes”, give the new arrangement.

- (v) For the new situation described in (iv) give an appropriate linear model and outline the ANOVA giving sources of variation and d.f.
- 10.4** A paint company wants to compare the abilities of four (4) white house paints to withstand environmental conditions. Eight (8) square houses, each with one side facing exactly north, are available in each of three states: Florida, Michigan, and California (that is, there are 24 houses altogether). Each side of a house is possibly exposed to different types of weather. Also, the houses are different from each other (because of different building materials, different ages, etc.). The company wants to paint each side of each house with a different paint. In addition to comparing the 4 paints the company is also interested in finding out whether differences among the paints vary from state to state.
- (i) Describe how you would set up the experiment, that is, what error-control design would you use. Explain the reasons for choosing the design.
 - (ii) Give an appropriate linear model for analyzing data from the experiment described in (i).
 - (iii) Outline the ANOVA table associated with the model given in (ii) (giving source of variation, d.f.) and indicate how you would investigate the questions the company is interested in.
 - (iv) Describe how you would perform the analysis in (iii) by using SAS or some other statistical package.
- 10.5** Derive the missing value formula (10.25).
- 10.6** Obtain the ANOVA table for the analysis of covariance for the Latin square design.
- 10.7** Derive expressions (10.26) and (10.27) for the estimated relative efficiencies of the Latin square design relative to the RCBD.
- 10.8** Extend model (10.30) to include replicate \times treatment interaction and obtain the ANOVA table for this model.
- 10.9** A Græco-Latin square in its pure form may not be very useful, but extension and replications of it often prove to be quite useful. Consider the following design:

Row	Column							
	1	2	3	4	5	6	7	8
1	$C\gamma$	$A\alpha$	$B\delta$	$A\beta$	$B\gamma$	$C\delta$	$D\alpha$	$D\beta$
2	$A\alpha$	$B\beta$	$D\beta$	$C\delta$	$A\delta$	$D\gamma$	$B\gamma$	$C\alpha$
3	$D\delta$	$C\gamma$	$A\gamma$	$B\alpha$	$D\alpha$	$A\beta$	$C\beta$	$B\delta$
4	$B\beta$	$D\delta$	$C\alpha$	$D\gamma$	$C\beta$	$B\alpha$	$A\delta$	$A\gamma$

where rows, columns and Greek letters are blocking factors.

- (i) What would you call this design? Give an appropriate linear model and outline the ANOVA table, giving sources of variation, d.f., and sums of squares.
- (ii) Give the SAS statements for analyzing data from such an experiment.
- (iii) Suppose you find out that for the experiment under consideration the “columns” represent animals. More specifically, columns 1–4 represent animals from one breed and columns 5–8 animals from a different breed. The researcher is interested in finding out whether differences among treatments are breed-specific.

Is the design given above appropriate for investigating this question? If yes, explain why; if no, indicate what you would have done differently.
- (iv) For the design in (iii) give an appropriate linear model and outline the ANOVA table, giving sources of variation and d.f.
- (v) Give the SAS statements for the analysis suggested in (iv).

- 10.10** Write out a linear model for the error-control design using the 4×4 completely orthogonalized square (Section 10.6.2) and obtain the ANOVA table for this design.
- 10.11** Suppose a $t \times t$ completely orthogonalized square is replicated r times. Write out one possible linear model for such an error-control design and obtain the associated ANOVA table.
- 10.12** Using the data of Example 10.15 show numerically that differences between treatment LS means do not depend on the choice of x in “if carry=‘0’ then carry=‘ x ’ ”.
- 10.13** Analyze the data of Example 10.15 without carry-over effects and compare the results with those of Table 10.11b.

This Page Intentionally Left Blank

CHAPTER 11

Factorial Experiments: Basic Ideas

11.1 INTRODUCTION

Our discussion so far has centered on error-reduction designs. As we have pointed out earlier (see Chapter 2), however, another component of experimental design is the treatment design. Here we shall be concerned especially with the situation where the treatments have a structure, more specifically, a factorial structure.

Suppose we have several factors, denoted by A, B, C, \dots . Each factor has a number of different expressions, or levels. For example, factor A may be “insecticide” and the levels represent different commercially available insecticides, labeled a_1, a_2, \dots, a_a ; factor B may be “amount of insecticide” with the levels representing the specific amounts, say $1, 2, 3, \dots$, units, generally denoted by b_1, b_2, \dots, b_b ; factor C may be “type of application” with the levels “manually” and “mechanically,” generally denoted by c_1, c_2, \dots, c_c . In this example the levels of factors A and C are qualitative, whereas the levels of factor B are quantitative. A treatment now consists of level combinations, one level from each factor, which we denote by $(a_i b_j c_k)$. These treatments are then applied in any of the error-control designs we have discussed earlier. The object is not so much to compare the treatments as such but to make statements about the “behavior” of the various factors, singly or jointly. We may ask, for example, “Is there a difference among the insecticides generally, that is, averaged over the levels of factors B and C ?”, or “Do the differences in the efficacies of the insecticides depend on the type of application?” The first question is one about the main effects of factor A , whereas the second question is concerned with the interaction between factors A and C . It is these types of questions and the fact that we can provide answers to them that make factorial experiments particularly valuable.

Factorial experiments can be used in various forms (for instance, Kempthorne, 1952): One procedure would be to estimate the effect of, say, factor A , keeping all the other factors at a constant level in one experiment; then estimate the effect of factor A after changing the level of factor B , and keeping the remaining factors at a constant

level in the next experiment; and so on. This procedure of varying one factor at a time would generally be used when the purpose is to establish a fundamental law as it would lead to detailed knowledge of the effect of one factor when the others are held constant. No information is, however, obtained on the dependence of the effects of a factor on the levels at which the other factors were held constant. To obtain such information we might use another experimental procedure, namely to vary the levels of each of the factors and consider all possible level combinations simultaneously. This would allow us to obtain information about main effects and, more importantly, about interactions among the various factors. These ideas and their practical applications in agronomic experimentation and more generally in scientific experimentation were introduced by Fisher (1935) and Yates (1937).

The value of factorial experiments lies in the fact that we look at several factors simultaneously which allows us to estimate the various effects and interactions and at the same time provides us with a wider inductive basis, that is, drawing conclusions over a wide range of conditions. Fisher (1935) has referred to this property as greater comprehensiveness. And even though, in general, we can estimate all possible interactions among factors, it is an empirical fact that the interactions among many factors (the so-called higher order interactions) are negligible for all practical purposes. This leads to a considerable reduction in the number of parameters, that is, main-effects and lower order interactions, and hence to an easier interpretation of data from a factorial experiment. It is for these reasons that factorial experiments are used widely in scientific and industrial experimentation. In the following sections we shall present some basic ideas about certain types of factorial experiments. A much more detailed and technical discussion will be given in Chapters II.7-16.

11.2 INFERENCES FROM FACTORIAL EXPERIMENTS

Suppose we have n factors A_1, A_2, \dots, A_n , where factor A_i has m_i levels $a_{i1}, a_{i2}, \dots, a_{im_i}$ ($i = 1, 2, \dots, n$). A treatment combination is denoted by $(a_{1i}a_{2j}a_{3k} \cdots a_{nl})$ and there are $\prod_{i=1}^n m_i$ such treatment combinations. We write the effect of a treatment combination as $\tau_{ijk \cdots l}$ and define the various main effects and interactions through an expansion of the type as illustrated for $n = 3$:

$$\begin{aligned} \tau_{ijk} = f(a_{1i}a_{2j}a_{3k}) = & \bar{\tau}_{...} + (\bar{\tau}_{i..} - \bar{\tau}_{...}) + (\bar{\tau}_{.j.} - \bar{\tau}_{...}) \\ & + (\bar{\tau}_{ij.} - \bar{\tau}_{i..} - \bar{\tau}_{.j.} + \bar{\tau}_{...}) + (\bar{\tau}_{.k.} - \bar{\tau}_{...}) \\ & + (\bar{\tau}_{i.k} - \bar{\tau}_{i..} - \bar{\tau}_{.k.} + \bar{\tau}_{...}) \\ & + (\bar{\tau}_{.jk} - \bar{\tau}_{.j.} - \bar{\tau}_{.k.} + \bar{\tau}_{...}) \\ & + (\tau_{ijk} - \bar{\tau}_{ij.} - \bar{\tau}_{i.k} - \bar{\tau}_{.jk} + \bar{\tau}_{i..} + \bar{\tau}_{.j.} + \bar{\tau}_{.k.} - \bar{\tau}_{...}). \quad (11.1) \end{aligned}$$

This is, of course, an identity in τ_{ijk} which gives rise to a model statement of the form

$$\begin{aligned} \tau_{ijk} = & \mu + A_{1i} + A_{2j} + (A_1A_2)_{ij} + A_{3k} \\ & + (A_1A_3)_{ik} + (A_2A_3)_{jk} + (A_1A_2A_3)_{ijk}, \quad (11.2) \end{aligned}$$

where μ represents the overall mean; A_{1i} , A_{2j} , A_{3k} represent the main effects associated with the factors A_1 , A_2 , A_3 ; $(A_1A_2)_{ij}$, $(A_1A_3)_{ik}$, $(A_2A_3)_{jk}$ represent the two-factor interactions associated with those factors, and $(A_1A_2A_3)_{ijk}$ represents the three-factor interaction. From (11.1) it follows immediately that

$$\sum_{i=1}^{m_1} A_{1i} = \sum_{j=1}^{m_2} A_{2j} = \sum_{k=1}^{m_3} A_{3k} = 0$$

$$\sum_{i=1}^{m_1} (A_1A_2)_{ij} = \sum_{j=1}^{m_2} (A_1A_2)_{ij} = 0, \text{ etc.}$$

and

$$\sum_{i=1}^{m_1} (A_1A_2A_3)_{ijk} = \sum_{j=1}^{m_2} (A_1A_2A_3)_{ijk} = \sum_{k=1}^{m_3} (A_1A_2A_3)_{ijk} = 0.$$

We have thus exploited the factorial structure of the treatments and decomposed the treatment effects into meaningful components. Of these the main effects and two-factor interactions (or first order interactions) are of major importance for the interpretation of data from such an experiment. The existence of higher order interactions of appreciable magnitude (relative to the main effects) makes the interpretation, unfortunately, much more difficult. As mentioned earlier, it is, however, an established fact that the importance, that is, magnitude, of higher order interactions tends to decrease as the number of factors involved increases (somewhat analogous to a Taylor series expansion). This fact will actually be exploited in later chapters (see also Chapters II.8-16) to construct useful incomplete block designs for factorial experiments.

Model (11.2) leads to a corresponding partitioning of the treatment sum of squares into main effect and interaction sums of squares as follows:

Source	d.f.
Treatments	$m_1m_2m_3 - 1$
A_1	$m_1 - 1$
A_2	$m_2 - 1$
$A_1 \times A_2$	$(m_1 - 1)(m_2 - 1)$
A_3	$m_3 - 1$
$A_1 \times A_3$	$(m_1 - 1)(m_3 - 1)$
$A_2 \times A_3$	$(m_2 - 1)(m_3 - 1)$
$A_1 \times A_2 \times A_3$	$(m_1 - 1)(m_2 - 1)(m_3 - 1)$

If the error-control design is an orthogonal design (CRD, RCBD, GRBD, LSD) then the various components in (11.2) are estimated by replacing in (11.1) the τ_{ijk} 's by the corresponding treatment means. The partitioning of the treatment sum of squares is then obtained by squaring each term on the right-hand side of (11.1), with the τ 's replaced by the treatment means, and summing each term over all subscripts. Tests of hypotheses are performed in the usual manner by using $MS(E)$ from the error-control design as the denominator in the F -statistic and the individual main effect or interaction sums of squares in the numerator.

For nonorthogonal designs (for example, incomplete block designs) we may set up a correspondence between ordinary treatments and factorial treatments to assign the treatment combinations to the various blocks. To estimate the main effects and interactions we replace in (11.1) the τ_{ijk} 's by the corresponding LS means. The (partial) sums of squares associated with main effects and interactions must be obtained by using the methods of Chapter 4, that is, by fitting full and reduced models.

Factorial experiments are most useful in exploratory work where the researcher is interested in investigating the effects of possibly a large number of factors over a certain range of levels and to find out whether the factors act additively, that is, independently, or whether they exhibit interaction. It is the broad picture here that is of primary interest and the researcher will have to use his subject-matter knowledge to select the treatment factors and determine their levels to be included in the experiment. Once a broad picture has been obtained then a more detailed study of factors judged to be important may be appropriate as a follow-up.

11.3 EXPERIMENTS WITH FACTORS AT TWO LEVELS

One of the disadvantages, from a practical point of view, of factorial experiments is the fact that the number of treatment combinations increases rapidly as the number of factors and/or levels increases. One way out of this dilemma is to consider only a subset of all possible treatment combinations, a so-called fractional factorial (see Section 11.5 and Chapters II.13, 14). Another possibility is to consider a reasonable number of factors and restrict for each factor the number of levels to 2. Those two levels may be chosen so that they cover, in some sense, the practical range of levels, whereas for other factors they represent the only possible levels. Suppose we have n such factors. Then we refer to this experiment as a 2^n factorial.

Although a 2^n factorial is commonly used we should emphasize that it is most useful as an exploratory experiment. This is particularly true if a factor admits more than two levels. If we restrict ourselves to two levels only, then we cannot examine the nature of the main effects and interactions in any detail, for example, in the case of quantitative factors we cannot examine trends other than linear, thus our earlier recommendation of follow-up studies of a smaller nature.

11.3.1 Definition of Main Effects and Interactions

We shall now consider briefly the definition of effects and interactions for a 2^n factorial as well as the estimation and testing of such effects. To keep the notation simple we shall illustrate the concepts for the special case $n = 3$. Extension to the general case should then be obvious.

Let us denote the three factors by A, B, C , and their levels by $a_0, a_1, b_0, b_1, c_0, c_1$, respectively. The eight treatment combinations can then be written (in standard or-

der) as

$$\begin{aligned}
 &a_0b_0c_0 \\
 &a_1b_0c_0 \\
 &a_0b_1c_0 \\
 &a_1b_1c_0 \\
 &a_0b_0c_1 \\
 &a_1b_0c_1 \\
 &a_0b_1c_1 \\
 &a_1b_1c_1.
 \end{aligned}$$

It is convenient to use the same notation also for the true response of those treatment combinations (from the context it should always be clear what is meant). We then define the following *simple effects* of A , denoted by $A(b_j, c_k)$, as the effect of factor A when changing A from level a_0 to level a_1 with factor B at level b_j and factor C at level c_k :

$$\begin{aligned}
 A(b_0, c_0) &= a_1b_0c_0 - a_0b_0c_0 \\
 A(b_1, c_0) &= a_1b_1c_0 - a_0b_1c_0 \\
 A(b_0, c_1) &= a_1b_0c_1 - a_0b_0c_1 \\
 A(b_1, c_1) &= a_1b_1c_1 - a_0b_1c_1.
 \end{aligned} \tag{11.3}$$

Using the definitions (11.3) we define the *main effect* A as

$$\begin{aligned}
 A &= \frac{1}{4} \sum_{j,k} A(b_j, c_k) \\
 &= \frac{1}{4} \left(\sum_{j,k} a_1b_jc_k - \sum_{j,k} a_0b_jc_k \right),
 \end{aligned} \tag{11.4}$$

that is, A represents the average change in response when a_0 is changed to a_1 . Symbolically, we express (11.4) as

$$A = \frac{1}{4}(a_1 - a_0)(b_1 + b_0)(c_1 + c_0), \tag{11.5}$$

where this expression is meaningful only when the right-hand side is multiplied out formally and the terms in that expression are interpreted as the true responses from the respective treatment combinations.

We can also define the effect of A when B is kept at level b_j and C is averaged over levels c_0 and c_1 as

$$\begin{aligned}
 A(b_j, \bar{c}) &= \frac{1}{2}[A(b_j, c_0) + A(b_j, c_1)] \\
 &= \frac{1}{2} \left[\sum_k a_1b_jc_k - \sum_k a_0b_jc_k \right]
 \end{aligned}$$

or, symbolically,

$$A(b_j, \bar{c}) = \frac{1}{2}(a_1 - a_0)b_j(c_1 + c_0) \tag{11.6}$$

for $j = 0, 1$. To the extent that $A(b_0, \bar{c})$ and $A(b_1, \bar{c})$ are different from each other, half the difference between them is defined as the *interaction* between the factors A and B , denoted by $A \times B$ or simply AB , that is,

$$AB = \frac{1}{2}[A(b_1, \bar{c}) - A(b_0, \bar{c})] \quad (11.7)$$

or symbolically,

$$AB = \frac{1}{4}(a_1 - a_0)(b_1 - b_0)(c_1 + c_0). \quad (11.8)$$

The factor $\frac{1}{2}$ in (11.7) is merely a convention so that the denominator in (11.8) represents, as in (11.5), the number of simple differences among treatment responses.

Note that AB as defined in (11.8) is the interaction between factors A and B averaged over the levels of factor C . We could, similar to (11.3) consider *simple interactions* between A and B , defined as

$$\begin{aligned} AB(c_0) &= \frac{1}{2}[A(b_1, c_0) - A(b_0, c_0)] \\ &= \frac{1}{2}(a_1 - a_0)(b_1 - b_0)c_0 \\ AB(c_1) &= \frac{1}{2}[A(b_1, c_1) - A(b_0, c_1)] \\ &= \frac{1}{2}(a_1 - a_0)(b_1 - b_0)c_1. \end{aligned}$$

The difference (apart from the factor $\frac{1}{2}$) between these two interactions is a measure of the *three-factor interaction* $A \times B \times C$, or simply ABC , that is,

$$\begin{aligned} ABC &= \frac{1}{2}[AB(c_1) - AB(c_0)] \\ &= \frac{1}{4}(a_1 - a_0)(b_1 - b_0)(c_1 - c_0). \end{aligned} \quad (11.9)$$

In a similar manner we can also define the main effects B and C , and the interactions AC and BC .

The reader can verify easily that each main effect and interaction represents a contrast among the treatment combinations and that these contrasts are orthogonal to each other. Apart from the factor $1/4$ the contrast coefficients are as given in Table 11.1. The reader will notice also that, for example, the coefficients for AB are the products of the corresponding coefficients for A and B , and so forth.

The general rule for writing down expressions like (11.5), (11.8), (11.9) and hence defining the main effects and interactions for the general 2^n factorial is as follows. Any effect or interaction X say, can be represented as

$$X = \frac{1}{2^{n-1}}(a_1 \pm a_0)(b_1 \pm b_0)(c_1 \pm c_0)(d_1 \pm d_0) \dots \quad (11.10)$$

where the sign in each bracket is positive if the corresponding capital letter is not contained in X and negative if it is contained in X , and the whole expression on the right-hand side of (11.10) is to be expanded algebraically and interpreted in terms of treatment combination responses. Just as illustrated in Table 11.1 for the 2^3 case, the main effects and interactions represent here, too, a set of $2^n - 1$ orthogonal contrasts.

Table 11.1 Contrast Coefficients for Main Effects and Interactions in 2³ Factorial

Main effect/ Interaction	$a_0b_0c_0$	$a_1b_0c_0$	$a_0b_1c_0$	$a_1b_1c_0$	$a_0b_0c_1$	$a_1b_0c_1$	$a_0b_1c_1$	$a_1b_1c_1$
<i>A</i>	-1	+1	-1	+1	-1	+1	-1	+1
<i>B</i>	-1	-1	+1	+1	-1	-1	+1	+1
<i>AB</i>	+1	-1	-1	+1	+1	-1	-1	+1
<i>C</i>	-1	-1	-1	-1	+1	+1	+1	+1
<i>AC</i>	+1	-1	+1	-1	-1	+1	-1	+1
<i>BC</i>	+1	+1	-1	-1	-1	-1	+1	+1
<i>ABC</i>	-1	+1	+1	-1	+1	-1	-1	+1

11.3.2 Estimation of Main Effects and Interactions

To estimate the main effects and interactions we first estimate the treatment effects from the error-reduction design used. In the case of orthogonal designs these are, of course, simply the treatment means, and this is the only case we shall discuss here. Let us denote for $n = 3$, the treatment mean for the treatment combination $(a_i b_j c_k)$ by $\bar{y}(a_i b_j c_k)$. In the expressions defining the main effects and interactions, such as (11.4) or (11.5), (11.8), (11.9), we then replace the true responses by the estimated responses, that is, the treatment means based on say r observations, where r is the number of replications in a CRD, $r = b$ is the number of blocks in a RCBD, $r = t$ is the number of rows and columns in a LSD, etc. We then obtain, for example,

$$\hat{A} = \frac{1}{4} [\bar{y}(a_1 b_0 c_0) + \bar{y}(a_1 b_1 c_0) + \bar{y}(a_1 b_0 c_1) + \bar{y}(a_1 b_1 c_1) - \bar{y}(a_0 b_0 c_0) - \bar{y}(a_0 b_1 c_0) - \bar{y}(a_0 b_0 c_1) - \bar{y}(a_0 b_1 c_1)]. \quad (11.11)$$

Assuming unit-treatment additivity in the broad sense and using the arguments as exposited in previous chapters we obtain immediately

$$\begin{aligned} \text{var}(\hat{A}) &= \frac{1}{16} \left[\sum_{j,k} \text{var}(\bar{y}(a_1 b_j c_k)) + \sum_{j,k} \text{var}(\bar{y}(a_0 b_j c_k)) \right] \\ &= \frac{1}{16} \left[8 \cdot \frac{1}{r} \right] \sigma_e^2 \\ &= \frac{1}{2r} \sigma_e^2. \end{aligned} \quad (11.12)$$

The other main effects and interactions are estimated analogously and each is estimated with variance given by (11.12).

These results are extended easily to the general case of n factors. Using (11.10) we then find

$$\hat{X} = \frac{1}{2^{n-1}} [\text{sum of } 2^{n-1} \text{ treatment means} - \text{sum of remaining } 2^{n-1} \text{ treatment means}] \quad (11.13)$$

and consequently,

$$\text{var}(\hat{X}) = \left(\frac{1}{2^{n-1}} \right)^2 \left[2^n \frac{1}{r} \right] \sigma_e^2 = \frac{1}{r 2^{n-2}} \sigma_e^2. \quad (11.14)$$

11.3.3 Sums of Squares for Main Effects and Interactions

Since each main effect and interaction represents a contrast among treatments, it is easy to obtain for the ANOVA the sum of squares associated with that contrast, say $\text{SS}(X)$. We know (see Chapter 7, equation (7.4)) that

$$\begin{aligned} \text{SS}(X) &= \frac{[\hat{X}]^2}{\text{var}(\hat{X})/\sigma_e^2} \\ &= r 2^{n-2} [\hat{X}]^2 \end{aligned} \quad (11.15)$$

using (11.14). Each $\text{SS}(X)$ has 1 d.f. and

$$E[\text{SS}(X)] = \sigma_e^2 + r 2^{n-2} [X]^2. \quad (11.16)$$

It is, of course, obvious then how the hypothesis $H_0: X = 0$ can be tested in the ANOVA.

The right-hand side of (11.16) also shows that if X is assumed to be negligible then $\text{SS}(X)$ may be pooled with $\text{SS}(E)$ to provide additional d.f. for error.

11.4 INTERPRETATION OF EFFECTS AND INTERACTIONS

The interpretation of effects and interactions follows closely from the definitions given in Section 11.3. For example, for the 2^3 factorial with factors A , B , and C , the main effect A is the effect of increasing factor A from the amount a_0 to the amount a_1 , averaging over all possible level combinations of factors B and C .

Now suppose we wish to obtain the effect of factor A , averaging over the low and high levels of factor B , that is, b_0, b_1 , but with factor C at the low level, that is, at level c_0 . Similar to (11.6) this effect is defined as

$$A(\bar{b}, c_0) = \frac{1}{2} (a_1 - a_0) (b_1 + b_0) c_0 = \frac{1}{2} [a_1 b_1 c_0 + a_1 b_0 c_0 - a_0 b_1 c_0 - a_0 b_0 c_0]. \quad (11.17)$$

From the definition of the main effect A , that is,

$$A = \frac{1}{4} (a_1 - a_0) (b_1 + b_0) (c_1 + c_0) \quad (11.18)$$

and the interaction AC , that is,

$$AC = \frac{1}{4} (a_1 - a_0) (b_1 + b_0) (c_1 - c_0) \quad (11.19)$$

it follows easily, by treating the right-hand sides of (11.18) and (11.19) as algebraic quantities, that $A(\bar{b}, c_0)$ of (11.17) can be expressed alternatively as

$$A(\bar{b}, c_0) = A - AC. \quad (11.20)$$

In a similar way we obtain the following:

$$A(\bar{b}, c_1) = A + AC$$

$$A(b_0, \bar{c}) = A - AB$$

$$A(b_1, \bar{c}) = A + AB$$

and for the simple effects defined in (11.3)

$$\begin{aligned} A(b_0, c_0) &= A - AB - AC + ABC \\ A(b_1, c_0) &= A + AB - AC - ABC \\ A(b_0, c_1) &= A - AB + AC - ABC \\ A(b_1, c_1) &= A + AB + AC + ABC. \end{aligned} \tag{11.21}$$

Algebraically, the expressions above can be written as

$$A(\bar{b}, c_0) = A(1 - C)$$

$$A(\bar{b}, c_1) = A(1 + C)$$

$$A(b_0, \bar{c}) = A(1 - B)$$

$$A(b_1, \bar{c}) = A(1 + B)$$

$$A(b_0, c_0) = A(1 - B)(1 - C)$$

$$A(b_1, c_0) = A(1 + B)(1 - C)$$

$$A(b_0, c_1) = A(1 - B)(1 + C)$$

$$A(b_1, c_1) = A(1 + B)(1 + C).$$

This gives an easy way of remembering them and of writing down the effect of any factor for any situation with regard to the other factors.

Another consequence of having expressions like (11.20) and (11.21) is that we can obtain easily estimates of these effects and the variances of these estimates. For example, we have

$$\hat{A}(\bar{b}, c_0) = \hat{A} - \hat{AC}.$$

Since A and AC are orthogonal contrasts among the treatment effects it follows that \hat{A} and \hat{AC} are uncorrelated and hence

$$\begin{aligned} \text{var}[\hat{A}(\bar{b}, c_0)] &= \text{var}(\hat{A}) + \text{var}(\hat{AC}) \\ &= 2 \cdot \frac{1}{r \cdot 2^{3-2}} \sigma_e^2 \\ &= \frac{1}{r} \sigma_e^2 \end{aligned}$$

and, similarly,

$$\hat{A}(b_0, c_0) = \hat{A} - \hat{AB} - \hat{AC} + \hat{ABC}$$

with

$$\text{var}[\hat{A}(b_0, c_0)] = 4 \frac{1}{r 2^{3-2}} \sigma_e^2 = \frac{2}{r} \sigma_e^2.$$

11.5 INTERACTIONS: A CASE STUDY

We have mentioned earlier that every treatment design needs to be imbedded in error-control design. We have also discussed the possibility of interaction between treatment and blocking factors (see Section 9.6.7), that is, between factors from the set \mathcal{X} and the sets \mathcal{Z} and \mathcal{U} (see Section 2.2.4). In this chapter we are considering the situation where the set \mathcal{X} consists of several factors, say $\mathcal{X} = \{A_1, A_2, \dots, A_n\}$. These factors give rise to main effects A_1, A_2, \dots, A_n and interactions of the form $A_i \times A_j (i, j = 1, 2, \dots, n; i \neq j)$, $A_i \times A_j \times A_k (i, j, k = 1, 2, \dots, n; i, j, k \text{ not equal, and so on})$. As a consequence, we now can also envision interactions of the form $A_i \times \mathcal{Z}$, $A_i \times \mathcal{U}$, $A_i \times A_j \times \mathcal{Z}$, $A_i \times A_j \times \mathcal{U}$. This list can be extended, of course, but as we have mentioned earlier, typically higher order interactions are negligible, or negligible from a practical point of view. Rather than discuss these possibilities in generalities we shall consider a particular experiment and point out some strategies for exploring the existence of the types of interactions mentioned above.

11.5.1 The Experiment

The following experiment was discussed by Pearce (1953, 1983) (see also Hinkelmann, 2004), but for the purpose of this discussion we have made slight modifications and have constructed the data (yield) based on summary data given in the article.

EXAMPLE 11.1: The objective is to study the effect of different pruning methods on the yield of varieties of pears. There are two treatment factors: $A_1 \equiv A = \text{type of pruning}$, $A_2 \equiv B = \text{amount of pruning}$, each with two levels. For factor A the two levels are: $F = \text{pruning with few leaders}$, $M = \text{pruning with many leaders}$, and for factor B the two levels are: $H = \text{hard pruning}$, $L = \text{light pruning}$. In order to broaden the scope of the study, the investigator included five varieties of pears: Am=Beurré d'Amanlis, Ha=Beurré Hardy, Co=Conference, Fe=Fertility, Pi=Pitmaston. These constitute the five levels of the intrinsic factor $z_1 = V = \text{variety}$. The experiment was set up as a randomized complete block design with six blocks for each variety (see Figure ??). Thus there is one non-specific factor $u_1 \equiv \beta = \text{block}$ with six levels (the original experiment had eight blocks for each variety).

Thus, in summary, the experiment is a 2^2 factorial experiment with treatments (F, H) , (M, H) , (F, L) , (M, L) in a randomized complete block design with a nested blocking structure $\beta(V)$ with $5 \times 6 = 30$ blocks of size four each. The four treatments were randomly assigned to four experimental units (trees) in each block (see an example in Figure 11.1).

11.5.2 The Model

Denoting the response to the treatment by y , we can write out a linear model analogous to (2.2) reflecting the treatment and block structures and the type of interactions mentioned above, as follows:

		Block					
		1	2	3	4	5	6
Am	(L. F)	(H. M)					
	(H. M)	(L. M)					
	(H. F)	(L. F)					
	(L. M)	(H. F)					
Ha							
Co							
Fe							
Pi							

Figure 11.1 Experimental Layout (Schematic).

$$\begin{aligned} y_{ijkl} = & \mu + V_i + \beta_{ij} + A_k + B_l + (AB)_{kl} \\ & + (VA)_{ik} + (VB)_{il} + (VAB)_{ikl} \\ & + (\beta A)_{ijk} + (\beta B)_{ijl} + (\beta AB)_{ijkl} + e_{ijkl}, \end{aligned} \tag{11.22}$$

where

- V_i = effect of i -th variety ($i = 1, 2, \dots, 5$)
- β_{ij} = effect of j -th block for i -th variety ($j = 1, 2, \dots, 5$)
- A_k = effect of k -th type of pruning ($k = 1(F), 2(M)$)

$B_l =$ effect of l -th amount of pruning ($l = 1(H), 2(L)$)

$(AB)_{kl} = A \times B$ interaction component

$(VA)_{ik} = A \times Z$ interaction component

$(VB)_{il} = B \times Z$ interaction component

$(VAB)_{ikl} = A \times B \times Z$ interaction component

$(\beta A)_{ijk} = A \times U$ interaction component

$(\beta B)_{ijl} = B \times U$ interaction component

$(\beta AB)_{ijkl} = A \times B \times U$ interaction component

Based on model (11.22) we can partition the total number of degrees of freedom, $119 = 120 - 1$, in the ANOVA table as given in Table 11.2. We note here that the effect terms contained in model (11.22) account for all the d.f., leaving no d.f. for error. We have done this on purpose and, in fact, encourage the reader to always write out what we might call a full model, that is, accounting for all possible effects and interactions and their associated d.f. This will provide a check whether in particular we have accounted for all interactions and what, if any, assumptions we need to make to obtain an adequate number of d.f. for error (in addition to possibly existing d.f. for pure error, such as exist for example in the GRBD (see Section 9.7).

11.5.3 The Analysis

We now consider the analysis of the data for the experiment described above. The data are given in Table 11.3. (The reader may notice that we have included a factor C, the meaning of which will be made clear in comment (iv) regarding Table 11.5).

Based on model (11.22) and the breakdown of the total d.f. we assume for the preliminary analysis that the interaction $AB\beta$ is negligible and hence used as the error term. We note that the d.f. associated with $A \times B \times \beta$ represent only part of the block \times treatment interaction d.f.

The analysis is performed using SAS PROC GLM. The input statements and the output are given in Table 11.4. We comment on the results as follows:

- (i) The $B \times \text{Block}(V)$ interaction is clearly non-significant ($P = .29$).
- (ii) The $A \times \text{Block}(V)$ interaction is most likely also negligible ($P = .16$).
- (iii) Based on the results in (i) and (ii) we may thus pool both interaction terms with the $A \times B \times \beta$ "interaction" to form the error term with 75 d.f. for future analysis purposes.
- (iv) Our new model then becomes

Table 11.2 ANOVA for Model (11.22)

Source of Variation	Degrees of Freedom
V	4
β	$25 = 5(6 - 1)$
A	1
B	1
$A \times B$	1
$V \times A$	4
$V \times B$	4
$V \times A \times B$	4
$\beta \times A$	25
$\beta \times B$	25
$\beta \times A \times B$	25
Total	119

$$y_{ijkl} = \mu + V_i + \beta_{ij} + A_k + B_l + (AB)_k \\ + (VA)_{ik} + (VB)_{il} + (VAB)_{ikl} + e_{ijkl}. \quad (11.23)$$

The SAS input statements for the ANOVA using model (11.23) and for some follow-up procedures are given in Table 11.5. Among these are the slice options 'LSMEANS $A * B$ /SLICE = B SLICE = A ' and 'LSMEANS $A * V$ $A * B * V$ /SLICE = V '. With regard to the $A * B$ interaction, the slice option tests whether the simple effects for A and B , respectively, are significant. In general, the slice option tests the equality of the LS means for one factor at the different levels of the other factor. With regard to the $V * A$ and $V * B$ interactions, the option 'SLICE = V ' enables us to test whether the simple effects of A and B are significant for each level of V . We note that we did include the option 'LSMEANS $A * B * V$ /SLICE = V ' only to show that this would result in testing whether the four LS means for (F, H) , (F, L) , (M, H) , and (M, L) are different from each other for every level of V , and that is of no interest to us.

We now turn to the analysis as presented in Table 11.5 and make the following comments:

- (i) The P -values for V and Block(V) should be ignored, since under randomization

Table 11.3 A Case Study (Data)

```

data pruning;
input V$ Block A$ B$ C Y@@;
datalines;
Am 1 F H 1 530 Am 1 F L 2 581 Am 1 M H 3 548 Am 1 M L 4 572
Am 2 F H 1 523 Am 2 F L 2 570 Am 2 M H 3 532 Am 2 M L 4 571
Am 3 F H 1 528 Am 3 F L 2 586 Am 3 M H 3 539 Am 3 M L 4 608
Am 4 F H 1 516 Am 4 F L 2 604 Am 4 M H 3 553 Am 4 M L 4 587
Am 5 F H 1 558 Am 5 F L 2 639 Am 5 M H 3 563 Am 5 M L 4 615
Am 6 F H 1 582 Am 6 F L 2 657 Am 6 M H 3 580 Am 6 M L 4 640
Ha 1 F H 1 534 Ha 1 F L 2 582 Ha 1 M H 3 554 Ha 1 M L 4 619
Ha 2 F H 1 538 Ha 2 F L 2 578 Ha 2 M H 3 543 Ha 2 M L 4 602
Ha 3 F H 1 563 Ha 3 F L 2 599 Ha 3 M H 3 567 Ha 3 M L 4 618
Ha 4 F H 1 567 Ha 4 F L 2 601 Ha 4 M H 3 601 Ha 4 M L 4 629
Ha 5 F H 1 547 Ha 5 F L 2 600 Ha 5 M H 3 607 Ha 5 M L 4 655
Ha 6 F H 1 582 Ha 6 F L 2 636 Ha 6 M H 3 602 Ha 6 M L 4 677
Co 1 F H 1 551 Co 1 F L 2 604 Co 1 M H 3 572 Co 1 M L 4 644
Co 2 F H 1 545 Co 2 F L 2 591 Co 2 M H 3 584 Co 2 M L 4 647
Co 3 F H 1 558 Co 3 F L 2 600 Co 3 M H 3 587 Co 3 M L 4 642
Co 4 F H 1 569 Co 4 F L 2 614 Co 4 M H 3 597 Co 4 M L 4 665
Co 5 F H 1 598 Co 5 F L 2 648 Co 5 M H 3 618 Co 5 M L 4 660
Co 6 F H 1 612 Co 6 F L 2 651 Co 6 M H 3 638 Co 6 M L 4 699
Fe 1 F H 1 575 Fe 1 F L 2 610 Fe 1 M H 3 590 Fe 1 M L 4 655
Fe 2 F H 1 554 Fe 2 F L 2 630 Fe 2 M H 3 605 Fe 2 M L 4 638
Fe 3 F H 1 576 Fe 3 F L 2 648 Fe 3 M H 3 608 Fe 3 M L 4 643
Fe 4 F H 1 595 Fe 4 F L 2 653 Fe 4 M H 3 631 Fe 4 M L 4 656
Fe 5 F H 1 609 Fe 5 F L 2 652 Fe 5 M H 3 641 Fe 5 M L 4 686
Fe 6 F H 1 597 Fe 6 F L 2 652 Fe 6 M H 3 660 Fe 6 M L 4 689
Pi 1 F H 1 600 Pi 1 F L 2 661 Pi 1 M H 3 625 Pi 1 M L 4 702
Pi 2 F H 1 606 Pi 2 F L 2 641 Pi 2 M H 3 635 Pi 2 M L 4 675
Pi 3 F H 1 610 Pi 3 F L 2 643 Pi 3 M H 3 642 Pi 3 M L 4 670
Pi 4 F H 1 609 Pi 4 F L 2 672 Pi 4 M H 3 653 Pi 4 M L 4 684
Pi 5 F H 1 632 Pi 5 F L 2 694 Pi 5 M H 3 669 Pi 5 M L 4 723
Pi 6 F H 1 655 Pi 6 F L 2 714 Pi 6 M H 3 676 Pi 6 M L 4 727
;
run;

```

Table 11.4 A Case Study (Preliminary ANOVA)

a) Input statements:

```
proc glm data=pruning;
class V Block A B C;
model Y=V Block(V) A B A*B V*A V*B V*A*B A*Block(V) B*Block(V);
run;
```

b) Output:

The GLM Procedure
Class Level Information

Class	Levels	Values
V	5	Am Co Fe Ha Pi
Block	6	1 2 3 4 5 6
A	2	F M
B	2	H L
C	4	1 2 3 4

Number of Observations Read 120
Number of Observations Used 120

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	94	261663.3833	2783.6530	30.58	<.0001
Error	25	2275.4167	91.0167		
Corrected Total	119	263938.8000			

R-Square 0.991379 Coeff Var 1.556578 Root MSE 9.540266 Y Mean 612.9000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
V	4	102743.0500	25685.7625	282.21	<.0001
Block(V)	25	50019.2500	2000.7700	21.98	<.0001
A	1	18451.2000	18451.2000	202.72	<.0001
B	1	78540.8333	78540.8333	862.93	<.0001
A*B	1	108.3000	108.3000	1.19	0.2858
V*A	4	3750.7167	937.6792	10.30	<.0001
V*B	4	306.5833	76.6458	0.84	0.5117
V*A*B	4	1494.7833	373.6958	4.11	0.0108
Block*A(V)	25	3415.5833	136.6233	1.50	0.1582
Block*B(V)	25	2833.0833	113.3233	1.25	0.2939

Table 11.5 A Case Study (ANOVA and Post-hoc Analysis)

a) Input statements:

```
proc glm data=pruning;
class V Block A B;
model Y=V Block(V) A B A*B A*V B*V A*B*V/SS3;
lsmeans A B A*B/slice=B slice=A;
estimate 'Main effect A' A -1 1;
estimate 'Main effect B' B -1 1;
lsmeans V A*V B*V A*B*V/slice=V;
run;
```

b) Output:

The GLM Procedure					
Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	44	255414.7167	5804.8799	51.07	<.0001
Error	75	8524.0833	113.6544		
Corrected Total	119	263938.8000			
	R-Square	Coeff Var	Root MSE	Y Mean	
	0.967704	1.739417	10.66088	612.9000	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
V	4	102743.0500	25685.7625	226.00	<.0001
Block (V)	25	50019.2500	2000.7700	17.60	<.0001
A	1	18451.2000	18451.2000	162.34	<.0001
B	1	78540.8333	78540.8333	691.05	<.0001
A*B	1	108.3000	108.3000	0.95	0.3321
V*A	4	3750.7167	937.6792	8.25	<.0001
V*B	4	306.5833	76.6458	0.67	0.6118
V*A*B	4	1494.7833	373.6958	3.29	0.0154
Least Squares Means					
	A	Y LSMEAN			
	F	600.500000			
	M	625.300000			

Table 11.5 (Continued)

Least Squares Means					
		B	Y LSMEAN		
		H	587.316667		
		L	638.483333		
		A	B	Y LSMEAN	
		F	H	573.966667	
		F	L	627.033333	
		M	H	600.666667	
		M	L	649.933333	
A*B Effect Sliced by B for Y					
B	DF	Sum of Squares	Mean Square	F Value	Pr > F
H	1	10693	10693	94.09	<.0001
L	1	7866.150000	7866.150000	69.21	<.0001
A*B Effect Sliced by A for Y					
A	DF	Sum of Squares	Mean Square	F Value	Pr > F
F	1	42241	42241	371.66	<.0001
M	1	36408	36408	320.34	<.0001
		V	Y LSMEAN		
		Am	574.250000		
		Co	612.250000		
		Fe	627.208333		
		Ha	591.708333		
		Pi	659.083333		
		V	A	Y LSMEAN	
		Am	F	572.833333	
		Am	M	575.666667	
		Co	F	595.083333	
		Co	M	629.416667	
		Fe	F	612.583333	
		Fe	M	641.833333	
		Ha	F	577.250000	
		Ha	M	606.166667	
		Pi	F	644.750000	
		Pi	M	673.416667	
V*A Effect Sliced by V for Y					
V	DF	Sum of Squares	Mean Square	F Value	Pr > F
Am	1	48.166667	48.166667	0.42	0.5170
Co	1	7072.666667	7072.666667	62.23	<.0001
Fe	1	5133.375000	5133.375000	45.17	<.0001
Ha	1	5017.041667	5017.041667	44.14	<.0001
Pi	1	4930.666667	4930.666667	43.38	<.0001

Table 11.5 (Continued)

V	B	Y LSMEAN
Am	H	546.000000
Am	L	602.500000
Co	H	585.750000
Co	L	638.750000
Fe	H	603.416667
Fe	L	651.000000
Ha	H	567.083333
Ha	L	616.333333
Pi	H	634.333333
Pi	L	683.833333

V*B Effect Sliced by V for Y

V	DF	Sum of Squares	Mean Square	F Value	Pr > F
Am	1	19154	19154	168.52	<.0001
Co	1	16854	16854	148.29	<.0001
Fe	1	13385	13385	119.53	<.0001
Ha	1	14553	14553	128.05	<.0001
Pi	1	14702	14702	129.35	<.0001

V	A	B	Y LSMEAN
Am	F	H	539.500000
Am	F	L	606.166667
Am	M	H	552.500000
Am	M	L	598.833333
Co	F	H	572.166667
Co	F	L	618.000000
Co	M	H	599.333333
Co	M	L	659.500000
Fe	F	H	584.333333
Fe	F	L	640.833333
Fe	M	H	622.500000
Fe	M	L	661.166667
Ha	F	H	555.166667
Ha	F	L	599.333333
Ha	M	H	579.000000
Ha	M	L	633.333333
Pi	F	H	618.666667
Pi	F	L	670.833333
Pi	M	H	650.000000
Pi	M	L	696.833333

V*A*B Effect Sliced by V for Y

V	DF	Sum of Squares	Mean Square	F Value	Pr > F
Am	3	19822	6607.277778	58.13	<.0001
Co	3	24235	8078.277778	71.08	<.0001
Fe	3	19195	6398.486111	56.30	<.0001
Ha	3	19725	6575.152778	57.85	<.0001
Pi	3	19675	6558.277778	57.70	<.0001

Dependent Variable: Y

Parameter	Estimate	Standard Error	t Value	Pr > t
Main effect A	24.8000000	1.94640219	12.74	<.0001
Main effect B	51.1666667	1.94640219	26.29	<.0001

theory no significance tests for block effects, that is, effects of intrinsic and non-specific factors, are permissible.

- (ii) The $A * B$ interaction is non-significant ($P = .33$). Thus there is no real need to invoke the slice option. We have included it, however, for purposes of illustration. Using the $A * B$ LS means and the definition of simple effects in Section 11.3.1 we find the estimates of the simple effects to be

$$\hat{A}(H) = 600.67 - 573.97 = 26.7$$

$$\hat{A}(L) = 649.93 - 627.03 = 22.9$$

and

$$\hat{B}(F) = 627.03 - 573.97 = 53.06$$

$$\hat{B}(M) = 649.93 - 600.67 = 49.26$$

and all are significantly different from zero ($P < .0001$) and so are the estimates of the main effects $\hat{A} = 24.8$ and $\hat{B} = 51.17$ with standard error 1.95. At the same time, the test for $A * B$ interaction indicates that the simple effects for A as well as those for B are not different from each other, leading to near-parallel lines in the interaction plot.

- (iii) The interactions $A * V$ and $A * B * V$ are significant ($P < .0001$) and $P = .015$, respectively). The results for the $A * V$ interaction sliced by V indicate that only the simple A -effects for variety Am are not different from each other, whereas the estimates of the simple A -effects for the other four varieties are of the same order of magnitude, around 30, as can be seen from the $A * V$ LS means. This interaction is clearly a codirectional interaction, and hence considering the overall A main effect is appropriate. Furthermore, the slice operation shows also that the $A \times V$ interaction comes about only because of the different behaviour of variety Am. If we were to drop Am from the analysis, there would be no $A \times V$ interaction. Since the $B \times V$ interaction is not significant, the $B * V$ slice operation is not really needed. The $B * V$ LS means show that the estimators of the simple B -effects are all about the same order of magnitude, around 50.
- (iv) Concerning the $A \times B \times V$ interaction, we have included the slice operation with $A * B * V$ effect sliced by V only to demonstrate that this operator does not produce the desired results for three-factor interactions. We would like to compare the simple $A \times B$ interactions (as defined in Section 11.3.1), but the results of the slice operation indicate that this procedure tests, for each variety, the equality to the four $A * B$ LS means, as indicated by $DF = 3$. To achieve our objective we now use the factor C introduced earlier, noting that the contrast vector $(1, -1, -1, 1)$ for C describes the $A \times B$ interaction. More specifically, we use the SAS input statement and the results in Table 11.6 as follows.

We note that from Tables 11.5 and 11.6 it follows that the following relationships among sums of squares exist:

$$SS(C) = SS(A) + SS(B) + SS(A * B)$$

Table 11.6 A Case Study (Additional Post-hoc Analysis)

a) Input statements:

```
proc glm data=pruning;
class V Block C;
model Y=V Block(V) C V*C/SS3;
lsmeans V*C;
estimate 'A*B for Am' C 1 -1 -1 1 V*C 1 -1 -1 1/divisor=2;
estimate 'A*B for Co' C 1 -1 -1 1 V*C 0 0 0 1 -1 -1 1/divisor=2;
estimate 'A*B for Fe' C 1 -1 -1 1 V*C 0 0 0 0 0 0 0 1 -1 -1
1/divisor=2;
estimate 'A*B for Ha' C 1 -1 -1 1 V*C 0 0 0 0 0 0 0 0 0 1 -1 -1
1/divisor=2;
estimate 'A*B for Pi' C 1 -1 -1 1 V*C 0 0 0 0 0 0 0 0 0 0 0 0 0 1
-1 -1 1/divisor=2;
run;
```

b) Output:

The GLM Procedure					
Dependent Variable: Y					
Source	DF	Squares	Sum of Mean Square	F Value	Pr > F
Model	44	255414.7167	5804.8799	51.07	<.0001
Error	75	8524.0833	113.6544		
Corrected Total	119	263938.8000			
	R-Square	Coeff Var	Root MSE	Y Mean	
	0.967704	1.739417	10.66088	612.9000	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
V	4	102743.0500	25685.7625	226.00	<.0001
Block(V)	25	50019.2500	2000.7700	17.60	<.0001
C	3	97100.3333	32366.7778	284.78	<.0001
V*C	12	5552.0833	462.6736	4.07	<.0001

Least Squares Means

V	C	Y LSMEAN
Am	1	539.500000
Am	2	606.166667
Am	3	552.500000
Am	4	598.833333
Co	1	572.166667
Co	2	618.000000
Co	3	599.333333
Co	4	659.500000
Fe	1	584.333333

Table 11.6 (Continued)

Fe	2	640.833333
Fe	3	622.500000
Fe	4	661.166667
Ha	1	555.166667
Ha	2	599.333333
Ha	3	579.000000
Ha	4	633.333333
Pi	1	618.666667
Pi	2	670.833333
Pi	3	650.000000
Pi	4	696.833333

Dependent Variable: Y

Parameter	Estimate	Standard Error	t Value	Pr > t
A*B for Am	-10.1666667	4.35228761	-2.34	0.0222
A*B for Co	7.1666667	4.35228761	1.65	0.1038
A*B for Fe	-8.9166667	4.35228761	-2.05	0.0440
A*B for Ha	5.0833333	4.35228761	1.17	0.2465
A*B for Pi	-2.6666667	4.35228761	-0.61	0.5419

and

$$SS(V * C) = SS(V * A) + SS(V * B) + SS(V * A * B)$$

with 3 and 12 d.f., respectively. From $SS(C) + SS(V * C)$ with 15 d.f. the estimate statements in Table 11.6a isolate 5 d.f. which specify the simple $A * B$ interactions for each variety. The results in Table 11.6 indicate that only the $A * B$ interaction for Am is clearly significant ($= .022$). The $A * B$ interaction for Fe is borderline significant ($P = .044$), whereas the other $A * B$ interactions are not significant. This shows again that the variety Am behaves somewhat differently than the other varieties. A closer look at the $V * A * B$ LS means confirms this finding as the highest yield for all varieties except Am is achieved for the (M , L) treatment combination. For Am the highest yield is obtained for the (F , L) treatment combination, but the difference between the yields for (M , L) and (F , L) is relatively small.

- (v) Thus, the overall conclusion from this study shows that over a wide range of pear varieties the method of light pruning with many leaders will lead to the best results. Possible exceptions are for varieties similar to Am, where light pruning with few leaders may produce slightly better results. \square

11.5.4 Separate Analyses

The follow-up procedures as described above are done within the context of the overall analysis using model 11.23. And this is the method we recommend in general. One

reason for proceeding in this way is that all inferences are based on the same error term, namely MS(Error) from the overall ANOVA, usually based on a sufficient number of degrees of freedom.

An alternative procedure, however, might be to perform separate analyses for each level of one intrinsic factor or for each level combination of several intrinsic factors. Since in our example the $V \times A \times B$ interaction is significant, we might be led to five analyses, each based on the model

$$Y_{jkl}^{(i)} = \mu^{(i)} + \beta_j^{(i)} + A_k^{(i)} + B_l^{(i)} + (AB)_{kl}^{(i)} + e_{jkl}^{(i)}$$

for $i = 1, 2, \dots, 5$. Although we would be able to make then recommendations separately for each variety, it becomes more difficult to arrive at statistically sound overall conclusions. We shall not provide the details of the five analyses here, but only report that the overall result would have been the same as obtained in (v) above.

11.5.5 Blocking by Intrinsic Factor Only

The experiment described and analyzed in the preceding sections uses an RCBD with a nested blocking structure and a factorial treatment structure. We have used these structures to investigate various forms of interactions. Generally speaking, we have considered interactions of the form $\mathcal{X} \times \mathcal{X} \times \mathcal{U}$, $\mathcal{X} \times \mathcal{U}$, $\mathcal{X} \times \mathcal{X} \times \mathcal{Z}$, $\mathcal{X} \times \mathcal{Z}$, $\mathcal{X} \times \mathcal{X}$. The absence of the $\mathcal{X} \times \mathcal{X} \times \mathcal{U}$ and $\mathcal{X} \times \mathcal{U}$ interactions provided us with an appropriate error term to perform the analysis in Section 11.5.3.

EXAMPLE 11.2:

In this section we shall consider the situation if the experiment had consisted of just one block for each variety. In that case “variety” is the only blocking factor. In other words, the intrinsic factor is the only blocking factor.

The typical approach to analyzing data from such an experiment would be to assume that the treatment \times block interaction is negligible and then to use the model

$$Y_{ijk} = \mu + V_i + A_j + B_k + (AB)_{jk} + e_{ijk}. \quad (11.24)$$

We know, however, from the analysis of the larger experiment that the $V \times A$ and $V \times A \times B$ interactions were significant there. Thus, the assumptions that lead to model (11.24) may not be appropriate.

In general, we do not have this kind of insight, but whenever an intrinsic factor is used as a blocking factor careful consideration must be given to possible existence of $\mathcal{X} \times \mathcal{Z}$ interactions. Usually such considerations have to be based on subject matter knowledge rather than on statistical arguments since there may only exist limited testing for the $\mathcal{X} \times \mathcal{Z}$ interaction as given by, for example, Tukey (1949) and Mandel (1961) (see Section 9.6).

In our example, the treatments have a factorial structure. Therefore the $\mathcal{X} \times \mathcal{Z}$ interaction can be divided into $A \times V$, $B \times V$, and $A \times B \times V$ interactions. This provides us with a choice whether to assume that all three interaction components or only one or two of them are negligible.

An obvious choice would be to assume that $A \times B \times V$ is negligible and to use $SS(A \times B \times V)$ as the error sum of squares, $SS(E)$, with 4 d.f. In our case we happen to know, however, that $A \times B \times V$ may not be negligible and hence we may not be willing to follow this route. Instead we shall propose here an ad-hoc approach to this potential problem.

11.5.6 Using the Half-normal Plot Technique

We shall adapt the method of half-normal plots which was proposed by Daniel (1959) to identify non-zero effects in a saturated fraction of a 2^n factorial. Saturated in this context means that the design does not provide any d.f. for error. This is the same situation here if we are not willing, a priori, to assume that some of the $\mathcal{X} \times \mathcal{Z}$ interactions are negligible. The method essentially consists of plotting the absolute values of the estimates of interactions and main effects with increasing magnitude on half-normal probability paper, and if the values are all zero they should lie on a straight line. Estimates with “large” deviations from this line are considered to be non-zero, that is, nonnegligible (for a description see Daniel (1959), Zahn (1975), and also Section II.13.9). In order to use this method for our purpose we need to partition the $\mathcal{X} \times \mathcal{Z}$ interaction into single-d.f.-contrasts. In general, if \mathcal{X} has ν_x d.f. and \mathcal{Z} has ν_z d.f., then $\mathcal{X} \times \mathcal{Z}$ has $\nu_x \cdot \nu_z$ d.f. Thus, there will be $\nu_x \cdot \nu_z$ contrasts which will have to be orthonormal, that is, orthogonal and normalized, for this procedure to work. We shall use our example to describe how this can be accomplished. The general procedure should then become obvious.

We have $\nu_x = 3$ and $\nu_z = 4$, where the 3 d.f. for \mathcal{X} are represented by those for A , B , and AB , and the 4 d.f. for \mathcal{Z} by four comparisons among the five varieties, denoted by $V1$, $V2$, $V3$, $V4$ say. For $V1$, $V2$, $V3$, $V4$ we choose the complete set of four orthogonal polynomials among the five varieties. The contrast coefficients for these orthogonal polynomials are given in Table 11.7 (see Section 7.4). We should note that these contrasts have no particular meaning here since the levels of V are nominal, but that they were chosen conveniently for mathematical purposes only; other contrasts could have been chosen just as well as long as they are orthogonal. The seven sets of contrast coefficients are given in Table 11.7, labelled $V1$, $V2$, $V3$, $V4$, A , B , AB . The coefficients for the 12 contrasts belonging to the $\mathcal{X} \times \mathcal{Z}$ interactions are then simply obtained by multiplying the coefficients for the corresponding \mathcal{X} and \mathcal{Z} contrasts. For example, the coefficients for the contrast $V1A$ is obtained by multiplying elementwise the coefficients for $V1$ and A . For the set of contrasts labelled $V1A$, $V2A$, \dots , $V4AB$ we also give the normalizing divisor (ND), which is the square root of the sum of the squared coefficients. We then obtain the contrast estimates and plot their absolute values on half-normal probability paper. The results, using for each variety the observations given in block 5, are given in Figure 11.2.

Inspection of Figure 11.2 shows that the absolute values for $V1A$, $V2A$, $V3A$, $V4A$, and $V2B$ do not lie on the line going through the smaller contrast values. This implies that, at least informally, these contrasts may not be negligible and, hence, probably should not be included in the error term.

Table 11.7 Single-d.f. Contrast Partition

V $A * B$	Λ_m			Λ_u			C_o			C_e			P_i			ND
	FH	FL	MH	FL	MH	ML	FH	FL	MH	FL	MH	ML	FH	FL	MH	
V_1	-2	-2	-2	-1	-1	-1	-1	-2	0	0	0	0	2	2	2	6.325
V_2	2	2	2	-1	-1	-1	-1	-2	-2	-2	-2	-2	2	2	2	7.483
V_3	-1	-1	-1	2	2	2	2	0	0	0	0	0	1	1	1	6.325
V_4	1	1	1	-4	-4	-4	-4	6	6	6	6	6	1	1	1	16.733
A	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	6.325
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	7.483
AB	1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	6.325
V_{1A}	-2	-2	2	-1	-1	-1	-1	0	0	0	0	0	2	2	2	16.733
V_{2A}	2	2	-2	-1	-1	-1	-1	-2	2	2	2	2	1	1	1	6.325
V_{3A}	-1	-1	1	2	2	2	2	0	0	0	0	0	1	1	1	6.325
V_{4A}	1	1	-1	-4	-4	-4	-4	6	6	6	6	6	-1	-1	-1	16.733
V_{1B}	-2	2	-2	-1	1	1	1	0	0	0	0	0	2	-2	2	6.325
V_{2B}	2	-2	2	-1	1	1	1	-2	-2	-2	-2	-2	1	-1	1	7.483
V_{3B}	-1	1	-1	2	-2	-2	-2	0	0	0	0	0	-1	1	-1	6.325
V_{4B}	1	-1	1	-4	4	4	4	6	-6	-6	-6	-6	1	-1	1	16.733
V_{1AB}	-2	2	2	1	1	1	1	0	0	0	0	0	2	-2	2	6.325
V_{2AB}	2	-2	-2	-1	-1	-1	-1	-2	2	2	2	2	1	1	1	7.483
V_{3AB}	-1	1	1	2	-2	-2	-2	0	0	0	0	0	-1	-1	-1	6.325
V_{4AB}	1	-1	-1	-4	4	4	4	6	-6	-6	-6	-6	1	-1	1	16.733

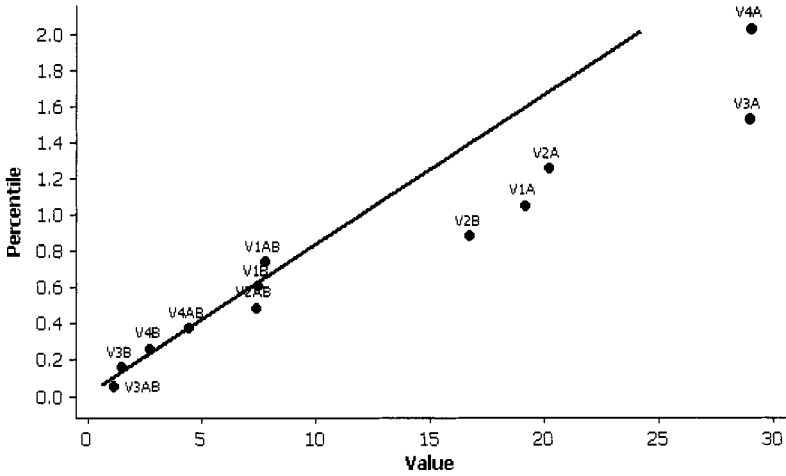


Figure 11.2 Half-normal Plot for $V \times A \times B$ Single-df Contrasts.

However, since the contrast $V2B$ represents only one d.f. of the 4 d.f. for the interaction $B \times V$, it may be quite appropriate from a practical point of view to declare $B \times V$ negligible and use the error term

$$SS(\text{Error}^*) = SS(B \times V) + SS(AB \times V). \quad (11.25)$$

The form of the error term (11.25) implies that the data should be analyzed according to the model

$$Y_{ijk} = \mu + V_i + A_j + B_k + (AB)_{jk} + (VA)_{ij} + e_{ijk}^*. \quad (11.26)$$

11.5.7 The Analysis

The analysis of the data using model 11.26 is presented in Table 11.8. We comment briefly on the SAS PROC GLM output:

- (i) The main effects A and B are significant with $P < .0001$.
- (ii) The interaction $A \times B$ is not significant ($P = .2023$).
- (iii) The $A \times V$ interaction is significant ($P = .0031$) as suggested already by the half-normal plot of Figure 11.2.

Table 11.8 Blocking by Variety Only

a) Input statements:

```
Data block5;
input V Block A B C Y@@;
datalines;

Am 5 F H 1 558 Am 5 F L 2 639 Am 5 M H 3 563 Am 5 M L 4 615
Ha 5 F H 1 547 Ha 5 F L 2 600 Ha 5 M H 3 607 Ha 5 M L 4 655
Co 5 F H 1 598 Co 5 F L 2 648 Co 5 M H 3 618 Co 5 M L 4 660
Fe 5 F H 1 609 Fe 5 F L 2 652 Fe 5 M H 3 641 Fe 5 M L 4 686
Pi 5 F H 1 632 Pi 5 F L 2 694 Pi 5 M H 3 669 Pi 5 M L 4 723
;
run;

proc glm data=block5;
class V A B;
model Y = V A B A*B V*A/SS3;
lsmeans A B A*B/stderr;
lsmeans V*A/stderr slice=V;
run;
```

b) Output:

The GLM Procedure					
Class Level Information					
Class	Levels	Values			
V	5	Am Co Fe Ha Pi			
A	2	F M			
B	2	H L			
Number of Observations Read			20		
Number of Observations Used			20		
Dependent Variable: Y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	39278.40000	3570.76364	59.79	<.0001
Error	8	477.80000	59.72500		
Corrected Total	19	39756.20000			
R-Square	Coeff Var	Root MSE	Y Mean		
0.987982	1.225336	7.728195	630.7000		

Table 11.8 (Continued)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
V	4	19287.70000	4821.92500	80.74	<.0001
A	1	3380.00000	3380.00000	56.59	<.0001
B	1	14045.00000	14045.00000	235.16	<.0001
A*B	1	115.20000	115.20000	1.93	0.2023
V*A	4	2450.50000	612.62500	10.26	0.0031

Least Squares Means

		Y LSMEAN	Standard Error	Pr > t
A				
F		617.700000	2.443870	<.0001
M		643.700000	2.443870	<.0001
B				
H		604.200000	2.443870	<.0001
L		657.200000	2.443870	<.0001
A B				
F	H	588.800000	3.456154	<.0001
F	L	646.600000	3.456154	<.0001
M	H	619.600000	3.456154	<.0001
M	L	667.800000	3.456154	<.0001
V A				
Am	F	598.500000	5.464659	<.0001
Am	M	589.000000	5.464659	<.0001
Co	F	623.000000	5.464659	<.0001
Co	M	639.000000	5.464659	<.0001
Fe	F	630.500000	5.464659	<.0001
Fe	M	663.500000	5.464659	<.0001
Ha	F	573.500000	5.464659	<.0001
Ha	M	631.000000	5.464659	<.0001
Pi	F	663.000000	5.464659	<.0001
Pi	M	696.000000	5.464659	<.0001

V*A Effect Sliced by V for Y

V	DF	Sum of Squares	Mean Square	F Value	Pr > F
Am	1	90.250000	90.250000	1.51	0.2539
Co	1	256.000000	256.000000	4.29	0.0722
Fe	1	1089.000000	1089.000000	18.23	0.0027
Ha	1	3306.250000	3306.250000	55.36	<.0001
Pi	1	1089.000000	1089.000000	18.23	0.0027

- (iv) Slicing the $A * V$ interaction indicates that the simple A -effects for variety Am are not significantly different from each other. Inspection of the $V * A$ LS means shows that the $A \times V$ interaction is codirectional. \square

11.5.8 Summary

It is virtually impossible to discuss all the possible ramifications that a complex treatment and blocking structure may have on the data model, assumptions about the terms in the model, the ensuing analysis, the implications of existing interactions, and the interpretation of the results. The two experiments discussed above serve as examples of some strategies we may apply.

First of all we suggest to write out a model including all possible interactions as determined by the treatment and blocking structure of a given design. Based on subject matter knowledge or previous results we may then decide to drop certain treatment-block factor interactions (we can **never** drop interactions between blocking factors, whether they are real or not), which then become part of the experimental error. In doing so we should always withstand the temptation of convenience or the desire to obtain additional d.f. for error. The latter may be accomplished through a preliminary test in the context of the ANOVA.

For the remaining interactions we have indicated different approaches, such as looking at LS means via interaction plots, using the SLICE operator in SAS PROC GLM, considering simple two-factor interactions and sets of orthogonal contrasts. Not always is it possible to arrive at a simple answer, especially if the structure is very complicated. The most important point is to stay within the objective of the experiment and perhaps formulate new objectives for a follow-up experiment.

11.6 2^n FACTORIALS IN INCOMPLETE BLOCKS

As mentioned earlier the number of treatment combinations in a factorial experiment may be quite large. If an error-reduction design with blocking has to be used we may not be in a position to have sufficiently homogeneous blocks large enough to accommodate all the treatment combinations. Hence some form of incomplete block design is called for. Obviously, any of the incomplete block designs we have described earlier can be used. For factorial experiments there exist, however, special methods of constructing incomplete block designs based on the assumption or knowledge that certain interactions are negligible or of lesser importance. We shall illustrate this for the 2^n factorial, specifically for the 2^3 factorial. Generalizations are discussed in detail in Chapters II.8-12.

11.6.1 2^3 Factorial in Blocks of Size 4

Suppose we have three factors A , B , C , and we have available blocks of size 4. If we assume that the three-factor interaction ABC is negligible we can then arrange the treatment combinations in such a way that ABC becomes nonestimable or, as we say, ABC

is *confounded with blocks*. The idea is to assign those treatment combinations which enter positively into ABC to one block and those that enter negatively into another block and then replicate this basic arrangement r times. Recall now (or see Table 11.1) that

$$\begin{aligned} ABC &= \frac{1}{4}(a_1 - a_0)(b_1 - b_0)(c_1 - c_0) \\ &= \frac{1}{4}[a_1b_1c_1 - a_1b_1c_0 - a_1b_0c_1 + a_1b_0c_0 - a_0b_1c_1 \\ &\quad + a_0b_1c_0 + a_0b_0c_1 - a_0b_0c_0]. \end{aligned}$$

Hence, the basic block arrangement is then as follows:

$$\begin{aligned} \text{Block 1: } & a_1b_1c_1, a_1b_0c_0, a_0b_1c_0, a_0b_0c_1 \\ \text{Block 2: } & a_1b_1c_0, a_1b_0c_1, a_0b_1c_1, a_0b_0c_0. \end{aligned} \quad (11.27)$$

Using the familiar model (suppressing subscripts)

$$y = \mu + \beta + \tau + e$$

for each observation, we see immediately that for our arrangement

$$E(\widehat{ABC}) = \beta_1 - \beta_2 + ABC. \quad (11.28)$$

This illustrates the phrase that ABC is nonestimable, namely that for this design \widehat{ABC} is a biased estimator for ABC , the bias being $\beta_1 - \beta_2$, the difference of block effects. Thus also the phrase: ABC is confounded with blocks, that is, ABC and $\beta_1 - \beta_2$ cannot be separated.

We see, however, from arrangement (11.27) that all other main effects and interactions are estimable in the usual way. The reason for that is that for every other effect each block contains two treatment combinations which enter positively into the effect and two which enter negatively so that the block effects cancel each other. Consider, for example, main effect A :

$$\begin{aligned} \text{Block 1: } & \text{positively: } a_1b_1c_1, a_1b_0c_0 \\ & \text{negatively: } a_0b_1c_0, a_0b_0c_1 \\ \text{Block 2: } & \text{positively: } a_1b_1c_0, a_1b_0c_1 \\ & \text{negatively: } a_0b_1c_1, a_0b_0c_0. \end{aligned}$$

Suppose we replicate the arrangement (11.27) r times, that is, we have $2r$ blocks of size 4 altogether. If we denote the block totals by B_i ($i = 1, 2, \dots, 2r$) and the grand total by G , we can then write the ANOVA as given in Table 11.9.

The important point to note here is that because of the confounding of ABC there are only 6 d.f. for treatments rather than the usual 7. This design is thus an example of what we have called earlier (see Section 9.8) a disconnected design, except that this is the result of a deliberate choice on our part.

11.6.2 2^3 Factorial in Blocks of Size 2

This method of constructing incomplete block arrangements or, as they are also called, *systems of confounding*, can be used for blocks with size equal to a power of 2 (the

Table 11.9 ANOVA for 2^3 Factorial in Blocks of Size 4 and ABC Confounded

Source	d.f.	SS	$E(\text{MS})$
Blocks	$2r - 1$	$\frac{1}{4} \sum_{i=1}^{2r} B_i^2 - \frac{1}{8r} G^2$	
Treatments	6		
<i>A</i>	1	$2r[\hat{A}]^2$	$\sigma_e^2 + 2r[A]^2$
<i>B</i>	1	$2r[\hat{B}]^2$	$\sigma_e^2 + 2r[B]^2$
<i>AB</i>	1	$2r[\hat{AB}]^2$	$\sigma_e^2 + 2r[AB]^2$
<i>C</i>	1	$2r[\hat{C}]^2$	$\sigma_e^2 + 2r[C]^2$
<i>AC</i>	1	$2r[\hat{AC}]^2$	$\sigma_e^2 + 2r[AC]^2$
<i>BC</i>	1	$2r[\hat{BC}]^2$	$\sigma_e^2 + 2r[BC]^2$
Error	$6(r - 1)$	Difference	σ_e^2
Total	$8r - 1$	$\sum y^2 - \frac{1}{8r} G^2$	

disadvantage, of course, is that it can only be used for blocks with size equal to a power of 2). For our example we shall now also consider the situation where we have available blocks of size 2.

The general idea is to first partition the treatment combinations into two sets based upon the sign with which they enter into a certain interaction, say ABC as above:

$$\begin{aligned} + : & \quad a_1b_1c_1, a_1b_0c_0, a_0b_1c_0, a_0b_0c_1 \\ - : & \quad a_1b_1c_0, a_1b_0c_1, a_0b_1c_1, a_0b_0c_0 \end{aligned}$$

Then each set above is partitioned again into sets of two based upon the sign with which they enter into another main effect or interaction, say BC . We then obtain the following partition:

Block	Sign for		Treatment Combination	
	ABC	BC		
1	+	+	$a_1b_1c_1, a_1b_0c_0$	(11.29)
2	+	−	$a_0b_1c_0, a_0b_0c_1$	
3	−	+	$a_0b_1c_1, a_0b_0c_0$	
4	−	−	$a_1b_1c_0, a_1b_0c_1$	

These four sets then constitute the basic arrangement in blocks of size 2, and this arrangement will be replicated r times, giving us $4r$ blocks.

The way in which arrangement (11.29) was constructed implies that ABC and BC are confounded with blocks. By inspection we also see that the main effect A is confounded with blocks, because

$$E(\hat{A}) = \frac{1}{2}(\beta_1 - \beta_2 - \beta_3 + \beta_4) + A$$

but all other effects are estimable, that is, not confounded. This is a consequence of our method of construction: Since we have four blocks with three d.f. among them, we have to confound three main effects or interactions (each with 1 d.f.) with blocks. Two of these three effects are chosen independently (ABC and BC in our case). The third effect is then determined automatically and can be obtained formally from

$$ABC \times BC = AB^2C^2 = A \quad (11.30)$$

that is, by formally multiplying the confounded effects into each other and then dropping any letter raised to the second power (for the mathematics behind this see Chapter II.7). We also refer to A in this case as the *generalized interaction* between ABC and BC . The general rule concerning confounding then says: if two effects are confounded with blocks, then their generalized interaction is also confounded with blocks.

11.6.3 Partial Confounding

It is generally undesirable to confound main effects and two-factor interactions with blocks. In our simple (but not unrealistic) example this is unavoidable. We can see this by simply listing all possible systems of confounding:

- (i) $ABC, AB, ABC \times AB = C$
- (ii) $ABC, AC, ABC \times AC = B$
- (iii) $ABC, BC, ABC \times BC = A$
- (iv) $AB, AC, AB \times AC = BC$ (11.31)
- (v) $A, B, A \times B = AB$
- (vi) $A, C, A \times C = AC$
- (vii) $B, C, B \times C = BC$.

Clearly, (v), (vi) and (vii) are the most undesirable systems, and only system (iv) avoids confounding main effects but at the price of confounding all three 2-factor interactions. To avoid complete loss of information about the effects in one of the sets in (11.31) (and thereby obtaining full information about the remaining effects), we may use a compromise solution by constructing a design based on several systems of confounding dictated largely by the requirements of the experiment. This may result in complete loss of information for some effects, partial information (to varying degrees) on other effects, and full information on the remaining effects. Such a method is referred to as *partial confounding* (as compared to *complete confounding* as described above). We shall now give a simple example.

EXAMPLE 11.3: Suppose the requirements of our experiment for the 2^3 factorial are

- (a) We would like as much information about main effects as possible.
- (b) All 2-factor interactions are equally important and we need some information about them.
- (c) The 3-factor interaction is most likely negligible.
- (d) We have available 16 blocks of size 2.

Since each of the systems of confounding listed in (11.31) gives rise to four blocks, one possibility then is to use four of those systems, each giving rise to a complete replicate. To satisfy the requirements of the experiment, we choose systems (i), (ii), (iii), (iv) and label the resulting arrangements Rep. I, Rep. II, Rep. III, Rep. IV, respectively, with the following confounded and estimable effects:

Replicate	Effects	
	Confounded	Estimable
I	C, AB, ABC	A, B, AC, BC
II	B, AC, ABC	A, C, AB, BC
III	A, BC, ABC	B, C, AB, AC
IV	AB, AC, BC	A, B, C, ABC

Over the whole experiment the amount of information for the individual effects is then as follows:

Effect	Amount of Information
A	$3/4$
B	$3/4$
AB	$1/2$
C	$3/4$
AC	$1/2$
BC	$1/2$
ABC	$1/4$

This seems to be a reasonable arrangement, in that it gives equal information about the main effects ($3/4$) and about the 2-factor interactions ($1/2$), and it gives some information about the 3-factor interaction.

The actual layout, that is, the assignment of the treatment combinations to the blocks can be obtained following the rule given above. The result is presented in Table 11.10.

As discussed above, all effects are estimable only from some of the replicates, namely those in which they are not confounded. This can be displayed as follows:

Effect	Estimable from Replicates	Effective # of replications (r)
A	I, II, IV	3
B	I, III, IV	3
AB	II, III	2
C	II, III, IV	3
AC	I, III	2
BC	I, II	2
ABC	IV	1

The difference in the effective number of replications implies that the effects are estimated with different precision, that is, in the general formula (11.14) for the variance of the estimate of an effect, r now takes on the values given above, that is,

$$\begin{aligned}\text{var}(\hat{A}) &= \text{var}(\hat{B}) = \text{var}(\hat{C}) = \frac{1}{3 \cdot 2} \sigma_e^2 \\ \text{var}(\widehat{AB}) &= \text{var}(\widehat{AC}) = \text{var}(\widehat{BC}) = \frac{1}{2 \cdot 2} \sigma_e^2 \\ \text{var}(\widehat{ABC}) &= \frac{1}{1 \cdot 2} \sigma_e^2.\end{aligned}$$

This unequal replication is, of course, also reflected in the ANOVA as given in Table 11.11. Here, for example, $\hat{A}_{\text{I,II,IV}}$ indicates that the main effect A is estimated from the observations in replicates I, II, IV only. \square

The examples discussed above are meant to be an introduction to the notion of confounding and partial confounding as well as the construction and analysis of appropriate designs. Using these examples the reader should have no difficulty applying these ideas to other 2^n factorials in appropriate incomplete blocks. The task will be made easier, however, by applying the mathematical tools provided in Chapters II.8 and 9. Extensions to other factorial experiments are discussed in detail in Chapters II.10–12. A convenient tool to generate systems of confounding is provided by SAS PROC FACTEX (SAS Institute, Inc. 2002 – 2003). For an illustration see Examples 11.11 and 11.12 in Section 11.11.

11.7 FRACTIONS OF 2^n FACTORIALS

11.7.1 Rationale for Fractional Replication

In our discussion so far we have used designs in which all treatment combinations have been used the same number of times. This may not always be very practical, in particular if the number of factors, n , is quite large. And as we have mentioned earlier, 2^n factorial experiments are very valuable for exploratory experiments involving a large number of factors.

Table 11.10 Partial Confounding of a 2^3 Factorial in Blocks of Size 2

Replicate	Block	Treatment Combinations
I	1	$a_1 b_1 c_1, a_0 b_0 c_1$
	2	$a_1 b_0 c_0, a_0 b_1 c_0$
	3	$a_1 b_1 c_0, a_0 b_0 c_0$
	4	$a_1 b_0 c_1, a_0 b_1 c_1$
II	5	$a_1 b_1 c_1, a_0 b_1 c_0$
	6	$a_1 b_0 c_0, a_0 b_0 c_1$
	7	$a_1 b_0 c_1, a_0 b_0 c_0$
	8	$a_1 b_1 c_0, a_0 b_1 c_1$
III	9	$a_1 b_1 c_1, a_1 b_0 c_0$
	10	$a_0 b_1 c_0, a_0 b_0 c_1$
	11	$a_0 b_1 c_1, a_0 b_0 c_0$
	12	$a_1 b_1 c_0, a_1 b_0 c_1$
IV	13	$a_0 b_0 c_1, a_1 b_1 c_0$
	14	$a_1 b_0 c_1, a_0 b_1 c_0$
	15	$a_1 b_1 c_1, a_0 b_0 c_0$
	16	$a_1 b_0 c_0, a_0 b_1 c_1$

The main reason for imposing the restriction that each of the treatment combinations is to be tested an equal number of times is that it results in the estimates of main effects and interactions having maximum precision and being uncorrelated. These, of course, are two reasonable and desirable properties. But is *maximum* precision really always necessary? Under what conditions can we achieve a precision that we may consider to be “*sufficient*” from a practical point of view?

The question we then ask here is whether it is always necessary to test all factorial combinations equally frequently or whether we can omit some of them. To get some insight into this question let us consider the following example.

EXAMPLE 11.4: Consider a 2^8 factorial, yielding 256 treatment combinations, but of the 255 d.f. only 36 account for the main effects (8) and the 2-factor interactions (28), with the remaining d.f. belonging to higher order interactions. Even if every treatment combination is tested only once, that is, $r = 1$, in blocks of size 16, say, and assuming that all interactions involving three or more factors are negligible, the breakdown of the d.f. in the ANOVA is as follows:

Table 11.11 ANOVA for Partially Confounded 2^3 Factorial of Table 11.10

Source	d.f.	SS	$E(MS)$
Blocks	15	$\frac{1}{2} \sum_{i=1}^{16} B_i^2 - \frac{G^2}{32}$	
Treatments	7		
A	1	$6[\widehat{A}_{I,II,IV}]^2$	$\sigma_e^2 + 6[A]^2$
B	1	$6[\widehat{B}_{I,III,IV}]^2$	$\sigma_e^2 + 6[B]^2$
AB	1	$4[\widehat{AB}_{II,III}]^2$	$\sigma_e^2 + 4[AB]^2$
C	1	$6[\widehat{C}_{II,III,IV}]^2$	$\sigma_e^2 + 6[C]^2$
AC	1	$4[\widehat{AC}_{I,III}]^2$	$\sigma_e^2 + 4[AC]^2$
BC	1	$4[\widehat{BC}_{I,II}]^2$	$\sigma_e^2 + 4[BC]^2$
ABC	1	$2[\widehat{ABC}_{IV}]^2$	$\sigma_e^2 + 2[ABC]^2$
Error	9	Difference	σ_e^2
Total	31	$\sum y^2 - \frac{G^2}{32}$	

Source	d.f.
Blocks	15
Main effects	8
2-factor interactions	28
Error	204
Total	255

The large number of d.f. for error (stemming from the negligible higher order interactions) may be more than is necessary and the precision that would result for the estimation of main effects and 2-factor interactions, given by a variance of $\sigma_e^2/64$, may well be unnecessarily high. We shall see later that under the assumptions made above 64 carefully chosen treatment combinations may indeed provide sufficient information. Such a design is called a *fractional factorial* or *fractional replication*, in this case a 1/4 replication of a 2^8 factorial. □

11.7.2 1/2 Fraction of the 2^3 Factorial

We shall now explain the concept of a fractional factorial, that is, the choice of the treatment combinations to be used, in terms of a simple (but not practical) example. Suppose we have $n = 3$ factors, A , B , C , and we can use only four treatment combinations. How should we choose them and, having chosen them, what kind of information can we obtain?

We are interested then in a $1/2$ fraction of the 2^3 factorial. If we assume that the interaction ABC is negligible, we might choose as the $1/2$ fraction either those treatment combinations which enter positively into ABC or those that enter negatively into ABC , that is,

$$\begin{aligned} + : & a_1b_1c_1, a_1b_0c_0, a_0b_1c_0, a_0b_0c_1 \\ - : & a_0b_0c_0, a_0b_1c_1, a_1b_0c_1, a_1b_1c_0. \end{aligned}$$

Suppose we choose the “+” fraction. Let us then examine how we would use these four treatment combinations to estimate main effects and 2-factor interactions. We can deduce this easily from the following table (which is obtained from Table 11.1):

	A	B	AB	C	AC	BC
$a_1b_1c_1$	+	+	+	+	+	+
$a_1b_0c_0$	+	–	–	–	–	+
$a_0b_1c_0$	–	+	–	–	+	–
$a_0b_0c_1$	–	–	+	+	–	–

The use of SAS PROC FACTEX to generate this design is illustrated in Example 11.13 (See Section 11.11). Substituting the treatment means $\bar{y}(a_ib_jc_k)$ for the treatment combinations $(a_ib_jc_k)$ we see, by inspection, that A is estimated in the same way as BC , B in the same way as AC , and C in the same way as AB ; that is,

$$E\left\{\frac{1}{2}[\bar{y}(a_1b_1c_1) + \bar{y}(a_1b_0c_0) - \bar{y}(a_0b_1c_0) - \bar{y}(a_0b_0c_1)]\right\} = A + BC \quad (11.32)$$

$$E\left\{\frac{1}{2}[\bar{y}(a_1b_1c_1) - \bar{y}(a_1b_0c_0) + \bar{y}(a_0b_1c_0) - \bar{y}(a_0b_0c_1)]\right\} = B + AC \quad (11.33)$$

$$E\left\{\frac{1}{2}[\bar{y}(a_1b_1c_1) - \bar{y}(a_1b_0c_0) - \bar{y}(a_0b_1c_0) + \bar{y}(a_0b_0c_1)]\right\} = C + AB \quad (11.34)$$

11.7.3 The Alias Structure

Equations (11.32)–(11.34) show that there is no way of estimating individually A , B , AB , etc., but only linear combinations of them. This can also be deduced intuitively by noticing that there are only three orthogonal contrasts among the four treatment combinations and those correspond to (11.32), (11.33), and (11.34). We then say that A and BC , B and AC , and C and AB are *confounded with each other* or *aliased*, that is, we cannot estimate them separately. Formally we can obtain the so-called *alias*

structure by realizing that the mean and ABC are confounded with each other which we express in the form of an algebraic identity as

$$I = ABC. \quad (11.35)$$

This relation is known as the *defining relation (contrast)* or the *identity relationship*. It determines the type of fraction we are choosing and the alias structure by interpreting I as unity and formally multiplying each effect into both sides of (11.35) and deleting any letter raised to the power 2; thus

$$A = A(ABC) = A^2BC = BC \quad (11.36)$$

$$B = B(ABC) = AB^2C = AC \quad (11.37)$$

$$C = C(ABC) = ABC^2 = AB. \quad (11.38)$$

Interpreting the equality sign as “confounded with,” then (11.32) and (11.36), (11.33) and (11.37), and (11.34) and (11.38) express the same results. In this sense relation (11.35) also means that ABC is confounded with the mean. If we want to indicate that we have chosen the “+” fraction we may write (11.35) more explicitly as

$$I = +ABC$$

or if we have chosen the (complementary) “−” fraction,

$$I = -ABC.$$

In that case, rather than estimating $A + BC$ we would be able to estimate $A - BC$, etc. The end result remains the same: the main effects are confounded with 2-factor interactions.

For this fraction, or other fractional factorials in general to be useful we must make additional assumptions. In our case we assume that all 2-factor interactions are negligible. Then all main effects become estimable. This $1/2$ fraction is therefore referred to as a *main effect plan*, also as a *resolution III design* (Box and Hunter, 1961), because the interaction (word) in (11.35) consists of three letters and as a consequence main effects are aliased with 2-factor interactions.

If instead of (11.35) we had chosen the defining relation

$$I = AB$$

to select a $1/2$ fraction, the alias structure would have been

$$A = B$$

$$C = ABC$$

$$AC = BC.$$

This seems a less desirable fraction, if only for the reason that two main effects are confounded.

This simple example has brought out several properties of fractional factorials:

- (i) Information is being “lost.”
- (ii) The fraction has to be chosen carefully to “minimize” the loss.
- (iii) Assumptions about higher order interactions have to be made in order to obtain (unbiased) estimates of main effects (and possibly low-order interactions).

11.7.4 1/4 Fraction of the 2^8 Factorial

It is clear from our discussion so far that the defining relation for a 1/2 fraction includes the highest-order interaction. But even a 1/2 fraction may still contain too many treatment combinations. Hence fractions of a high degree, such as 1/4, 1/8, etc., may have to be considered as viable designs. To illustrate this we now return to Example 11.4, a 1/4 fraction of a 2^8 factorial with factors A, B, C, D, E, F, G, H .

In principle we can proceed as follows:

- (i) divide the set of the 2^8 treatment combinations into two sets based upon the sign with which they enter into a chosen interaction, E_1 say;
- (ii) choose one of those two sets;
- (iii) divide the chosen set again into two sets based upon the sign with which the treatment combinations enter into another designated interaction, E_2 , say.

Since interactions are orthogonal contrasts we know that this will result in a set of $2^6 = 64$ treatment combinations. However, just as in constructing systems of confounding (Section 11.6) we must be careful in our choice of E_1 and E_2 for the following reason. Since all 64 chosen treatment combinations have the same sign in E_1 and the same sign in E_2 , E_1 and E_2 are confounded with the mean. It is easy to see, however, that the generalized interaction $E_3 = E_1 E_2$ is then also confounded with the mean, that is, the 64 chosen treatment combinations also have the same sign in E_3 (see Chapter II.13). The question then is: How should we choose E_1 and E_2 and hence E_3 to obtain a fraction with the “most reasonable” alias structure knowing that this will be determined from the defining relation

$$I = E_1 = E_2 = E_3. \quad (11.39)$$

An intuitive approach might be to start with the highest-order interaction for E_1 and some other interaction for E_2 , but this may lead to a low-order interaction for E_3 and hence to an undesirable alias structure in that effects which we would like to estimate are confounded with each other. In order to approach this problem more systematically, we first have to decide which effects we want to estimate and which interactions we may assume to be negligible. Suppose we want to estimate (if possible) all main effects and 2-factor interactions and we assume that all other interactions are negligible. This means that main effects and 2-factor interactions cannot be confounded with other main effects and/or 2-factor interactions. To be assured of this we must have in (11.39) that E_1 , E_2 and E_3 are at least 5-factor interactions. Suppose then we choose

$E_1 = ABCDE$, $E_2 = ABFGH$, then $E_3 = (ABCDE)(ABFGH) = CDEFGH$, which obviously is of the required form. Hence we have

$$I = ABCDE = ABFGH = CDEFGH. \quad (11.40)$$

This defining relation indicates that main effects will be confounded with interactions involving four or more factors, and 2-factor interactions are confounded with interactions involving three or more factors, for instance,

$$A = BCDE = BFGH = ACDEFGH$$

$$AC = BDE = BCFGH = ADEFGH.$$

In addition, (11.40) determines the 1/4 fraction. Suppose we decide to choose all treatment combinations which enter negatively into $ABCDE$ and $ABFGH$, we see from the definitions

$$ABCDE = \frac{1}{2^7}(a_1 - a_0)(b_1 - b_0)(c_1 - c_0)(d_1 - d_0)$$

$$(e_1 - e_0)(f_1 + f_0)(g_1 + g_0)(h_1 + h_0)$$

$$ABFGH = \frac{1}{2^7}(a_1 - a_0)(b_1 - b_0)(c_1 + c_0)(d_1 + d_0)$$

$$(e_1 + e_0)(f_1 - f_0)(g_1 - g_0)(h_1 - h_0)$$

that the selected treatment combinations must have an odd number of factors from A, B, C, D, E and A, B, F, G, H at the low (zero) level. Those treatment combinations are given in Table 11.12 (in order to simplify the notation we write a treatment combination as (x_1, x_2, \dots, x_8) where $x_i = 0, 1$ ($i = 1, 2, \dots, 8$) indicating the low and high level of the i th factor, respectively).

This fractional factorial is also called a *resolution V design* since the lowest-order interaction contained in (11.40) has five factors (letters) and consequently main effects are aliased with 4-factor interactions and 2-factor interactions are aliased with 3-factor interactions. In many situations a resolution V design is the most desirable fraction since it allows the estimation of all main effects and 2-factor interactions, assuming that all other interactions are negligible. But even such a fractional factorial may be too large, hence the need for resolution III and resolution IV designs. *Resolution IV designs* are fractions in which main effects are confounded with 3-factor interactions and 2-factor interactions are confounded with other 2-factor interactions (see Section II.13.3.2).

11.7.5 Systems of Confounding for Fractional Factorials

As mentioned earlier it may not be possible to arrange the 64 treatment combinations in a CRD with $r = 1$. Instead we may consider, an incomplete block design with $b = 4$ blocks and $k = 16$ EUs per block. To do so we make use of the method described in Section 11.6.

Table 11.12 1/4 Fraction of the 2^8 Factorial

T.C.									T.C.								
#	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	#	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1.	0	0	0	0	0	0	0	0	33.	1	0	1	1	1	0	0	0
2.	1	1	0	0	0	0	0	0	34.	0	1	1	1	1	0	0	0
3.	0	0	1	1	0	0	0	0	35.	1	0	0	0	1	0	0	0
4.	1	1	1	1	0	0	0	0	36.	0	1	0	0	1	0	0	0
5.	0	0	1	0	1	0	0	0	37.	1	0	0	1	0	0	0	0
6.	1	1	1	0	1	0	0	0	38.	0	1	0	1	0	0	0	0
7.	0	0	0	1	1	0	0	0	39.	1	0	1	0	0	0	0	0
8.	1	1	0	1	1	0	0	0	40.	0	1	1	0	0	0	0	0
9.	0	0	0	0	0	1	1	0	41.	1	0	1	1	1	1	1	0
10.	1	1	0	0	0	1	1	0	42.	0	1	1	1	1	1	1	0
11.	0	0	1	1	0	1	1	0	43.	1	0	0	0	1	1	1	0
12.	1	1	1	1	0	1	1	0	44.	0	1	0	0	1	1	1	0
13.	0	0	1	0	1	1	1	0	45.	1	0	0	1	0	1	1	0
14.	1	1	1	0	1	1	1	0	46.	0	1	0	1	0	1	1	0
15.	0	0	0	1	1	1	1	0	47.	1	0	1	0	0	1	1	0
16.	1	1	0	1	1	1	1	0	48.	0	1	1	0	0	1	1	0
17.	0	0	0	0	0	1	0	1	49.	1	0	1	1	1	1	0	1
18.	1	1	0	0	0	1	0	1	50.	0	1	1	1	1	1	0	1
19.	0	0	1	1	0	1	0	1	51.	1	0	0	0	1	1	0	1
20.	1	1	1	1	0	1	0	1	52.	0	1	0	0	1	1	0	1
21.	0	0	1	0	1	1	0	1	53.	1	0	0	1	0	1	0	1
22.	1	1	1	0	1	1	0	1	54.	0	1	0	1	0	1	0	1
23.	0	0	0	1	1	1	0	1	55.	1	0	1	0	0	1	0	1
24.	1	1	0	1	1	1	0	1	56.	0	1	1	0	0	1	0	1
25.	0	0	0	0	0	0	1	1	57.	1	0	1	1	1	0	1	1
26.	1	1	0	0	0	0	1	1	58.	0	1	1	1	1	0	1	1
27.	0	0	1	1	0	0	1	1	59.	1	0	0	0	1	0	1	1
28.	1	1	1	1	0	0	1	1	60.	0	1	0	0	1	0	1	1
29.	0	0	1	0	1	0	1	1	61.	1	0	0	1	0	0	1	1
30.	1	1	1	0	1	0	1	1	62.	0	1	0	1	0	0	1	1
31.	0	0	0	1	1	0	1	1	63.	1	0	1	0	0	0	1	1
32.	1	1	0	1	1	0	1	1	64.	0	1	1	0	0	0	1	1

EXAMPLE 11.5: For a $\frac{1}{4} = \frac{1}{2^2}$ fraction of a 2^8 factorial, that is, a 2^{8-2} fractional factorial we obviously do not want to confound main effects and 2-factor interactions. These account for only 36 estimable functions of effects out of the 63 available. We therefore need to find two estimable functions of effects which were assumed to be negligible earlier. We then confound these effects and their generalized interaction with blocks. It follows from (11.40) that, for example,

$$ACF = BDEF = BCGH = ADEGH \quad (11.41)$$

and

$$BDG = ACEG = ADFH = BCEFH \quad (11.42)$$

are such functions and their generalized interaction

$$ABCDGF = EFG = CDH = ABEH \quad (11.43)$$

is also of such form. The four blocks are then obtained by considering the signs with which each of the 64 treatment combinations obtained from (11.40) and given in Table 11.12 enters into ACF (11.41) and BDG (11.42) as follows:

Block	Sign for	
	ACF	BDG
1	+	+
2	+	−
3	−	+
4	−	−

In Table 11.13 we give those signs for each treatment combination and in Table 11.14 we give the final design.

The main effects and 2-factor interactions can then be estimated in the usual way. For an effect X we have

$$\hat{X} = \frac{1}{2^{8-2-1}} [\text{sum of } 2^5 \text{ obs.} - \text{sum of remaining } 2^5 \text{ obs.}]$$

with

$$\text{var}(\hat{X}) = \frac{1}{2^{8-2-2}} \sigma_e^2 = \frac{1}{2^4} \sigma_e^2.$$

We also have for the ANOVA (as outlined earlier)

$$\text{SS}(X) = 2^4 [\hat{X}]^2.$$

The sums of squares associated with higher order interactions, except those given in (11.41), (11.42), and (11.43), are, of course, part of the $\text{SS}(\text{Error})$. \square

The methods of obtaining the $1/4$ fraction and the block arrangements may appear rather tedious, but they illustrate the underlying principles. More expeditious methods are described in Chapters II.13 and 14 together with methods of obtaining fractional factorials for other factorial experiments. See also Section 11.11 illustrating SAS PROC FACTEX (SAS Institute, Inc. 2002 – 2003).

Table 11.13 Signs with Which Treatment Combinations
in Table 11.12 Enter into *ACF* and *BDG*

T.C. #	<i>ACF</i>	<i>BDG</i>	T.C. #	<i>ACF</i>	<i>BDG</i>
1.	—	—	33.	—	+
2.	+	+	34.	+	—
3.	+	+	35.	+	—
4.	—	—	36.	—	+
5.	+	—	37.	+	+
6.	—	+	38.	—	—
7.	—	+	39.	—	—
8.	+	—	40.	+	+
9.	+	+	41.	+	—
10.	—	—	42.	—	+
11.	—	—	43.	—	+
12.	+	+	44.	+	—
13.	—	+	45.	—	—
14.	+	—	46.	+	+
15.	+	—	47.	+	+
16.	—	+	48.	—	—
17.	+	—	49.	+	+
18.	—	+	50.	—	—
19.	—	+	51.	—	—
20.	+	—	52.	+	+
21.	—	—	53.	—	+
22.	+	+	54.	+	—
23.	+	+	55.	+	—
24.	—	—	56.	—	+
25.	—	+	57.	—	—
26.	+	—	58.	+	+
27.	+	—	59.	+	+
28.	—	+	60.	—	—
29.	+	+	61.	+	—
30.	—	—	62.	—	+
31.	—	—	63.	—	+
32.	+	+	64.	+	—

Table 11.14 1/4 Fraction of 2^8 Factorial in Blocks of Size 16

Block 1								Block 2							
1	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0
0	0	1	1	0	0	0	0	1	1	0	1	1	0	0	0
0	0	0	0	0	1	1	0	1	1	1	0	1	1	1	0
1	1	1	1	0	1	1	0	0	0	0	1	1	1	1	0
1	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1
0	0	0	1	1	1	0	1	1	1	1	1	0	1	0	1
0	0	1	0	1	0	1	1	1	1	0	0	0	0	1	1
1	1	0	1	1	0	1	1	0	0	1	1	0	0	1	1
1	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0
0	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0
0	1	0	1	0	1	1	0	1	0	1	1	1	1	1	0
1	0	1	0	0	1	1	0	0	1	0	0	1	1	1	0
1	0	1	1	1	1	0	1	0	1	0	1	0	1	0	1
0	1	0	0	1	1	0	1	1	0	1	0	0	1	0	1
0	1	1	1	1	0	1	1	1	0	0	1	0	0	1	1
1	0	0	0	1	0	1	1	0	1	1	0	0	0	1	1
Block 3								Block 4							
1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	1	1	1	1	0	0	0	0
0	0	1	0	1	1	1	0	1	1	0	0	0	1	1	0
1	1	0	1	1	1	1	0	0	0	1	1	0	1	1	0
1	1	0	0	0	1	0	1	0	0	1	0	1	1	0	1
0	0	1	1	0	1	0	1	1	1	0	1	1	1	0	1
0	0	0	0	0	0	1	1	1	1	1	0	1	0	1	1
1	1	1	1	0	0	1	1	0	0	0	1	1	0	1	1
1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0
0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
0	1	1	1	1	1	1	0	1	0	0	1	0	1	1	0
1	0	0	0	1	1	1	0	0	1	1	0	0	1	1	0
1	0	0	1	0	1	0	1	0	1	1	1	1	1	0	1
0	1	1	0	0	1	0	1	1	0	0	0	1	1	0	1
0	1	0	1	0	0	1	1	1	0	1	1	1	0	1	1
1	0	1	0	0	0	1	1	0	1	0	0	1	0	1	1

11.8 ORTHOGONAL MAIN EFFECT PLANS FOR 2^n FACTORIALS

Among the fractional factorials discussed in the previous section, resolution III designs or main effect plans are of particular importance. We shall give a brief discussion here but defer details to Chapter II.14.

As mentioned earlier, the value of factorial experiments in general lies in the fact that higher order interactions are usually negligible. This leads to a considerable reduction in the number of parameters, that is, treatment effects or, more specifically, treatment effect contrasts, that need to be considered in the analysis of the data from such experiments. This, in turn, also leads to a reduction in the number of treatment combinations to be used in an experiment and hence to a reduction in the number of observations to be taken. It is in this sense that factorial experiments can be very economical.

The extreme situation is obviously achieved if all interactions can be considered negligible, so that the treatment effects can be represented in terms of main effects only.

EXAMPLE 11.6: For $n = 3$, we can rewrite model (11.2) as

$$\tau_{ijk} = \mu + A_{1i} + A_{2j} + A_{3k}. \quad (11.44)$$

For the 2^3 situation, which is of concern here, (11.44) can be rewritten in terms of the main effects as defined in (11.5) and Table 11.1 as follows (replacing A_1 by A , A_2 by B , and A_3 by C):

$$\tau_{ijk} = \mu \pm \frac{1}{2}A \pm \frac{1}{2}B \pm \frac{1}{2}C \quad (11.45)$$

where $i, j, k = 0, 1$ and the signs on the right-hand side of (11.45) depend on the values of i, j , and k in that the minus sign is used if the corresponding subscript is 0 and the plus sign is used if the corresponding subscript is 1. For example,

$$\tau_{011} = a_0b_1c_1 = \mu - \frac{1}{2}A + \frac{1}{2}B + \frac{1}{2}C$$

The equivalence of (11.44) and (11.45) can be verified easily by using the definition of the main effects for the 2^n factorial and taking $\mu = \sum a_i b_j c_k / 2^3$ (for details see Chapter II.7). The model (11.45) can obviously be extended to the general 2^n factorial. \square

In general then, if the assumption of no interactions is reasonable, we need to be able to estimate only $1 + n$ parameters (the mean and n main effects) for an experiment with n factors each at two levels. We have seen in Section 11.6.2 that, for example, for $n = 3$ this can be achieved with 4 treatment combinations, that is, a $\frac{1}{2}$ -fraction of the 2^3 experiment. The treatment combinations (runs) used are listed below where the levels of the factors A, B, C for runs 1, 2, 3, 4 are given in the body of the table:

Run #	Factor			Observation
	A	B	C	
1	1	1	1	y_1
2	1	0	0	y_2
3	0	1	0	y_3
4	0	0	1	y_4

Inspection of the table shows that for each pair of factors, that is, (A, B) , (A, C) , (B, C) , each possible ordered combination of zeros and ones occurs the same number of times, in this case once. Such arrangements are called *orthogonal arrays* (more precisely, orthogonal arrays of strength two) and in connection with fractional factorials the design is called an *orthogonal main effect plan*. As the name suggests these plans allow the estimation of main effects (under the assumption of no interactions), and the estimators for the main effects are uncorrelated. For example, with the observations for the design above denoted by y_1, y_2, y_3, y_4 , we have

$$\hat{A} = \frac{1}{2}(y_1 + y_2 - y_3 - y_4)$$

and

$$\hat{B} = \frac{1}{2}(y_1 - y_2 + y_3 - y_4)$$

with

$$\text{cov}(\hat{A}, \hat{B}) = \frac{1}{4}(\sigma_e^2 - \sigma_e^2 - \sigma_e^2 + \sigma_e^2) = 0$$

and so on.

The main effect plans described above have been given considerable prominence in industrial and process development, associated with the name Taguchi (for example, Taguchi, 1986). An interesting example of applying a main effect plan in product development is a tile experiment described by Taguchi (1986, pp. 80–83):

EXAMPLE 11.7: Seven factors, all related to the apportionment of materials in tile production and each having two levels (level 0 = level used in current production, level 1 = level thought to be superior in terms of cost and quality), are to be investigated in an effort to find the “best” combination of levels. More precisely, the factors and their levels were the following:

Factor	Level	
	0	1
A: Lime additive content	1%	5%
B: Granularity of additive	coarse	fine
C: Agalmatolite content	53%	43%
D: Type of agalmatolite	current mixture	less expensive mixture
E: Charge quantity	1,300 kg	1,200 kg
F: Waste return content	4%	0%
G: Feldspar content	5%	0%

The design used was an orthogonal main effect plan for 7 factors using the following 8 treatment combinations:

Run #	Factor						
	A	B	C	D	E	F	G
1	0	0	0	0	0	0	0
2	0	0	0	1	1	1	1
3	0	1	1	0	0	1	1
4	0	1	1	1	1	0	0
5	1	0	1	0	1	0	1
6	1	0	1	1	0	1	0
7	1	1	0	0	1	1	0
8	1	1	0	1	0	0	1

The data consisted of the percent defective tiles. After estimating the main effects and using a model of the form (11.45) the optimum set of conditions was found to be $a_1b_1c_0d_0e_1f_1g_0$. (Taguchi uses a slightly different argument and the reader is referred to his account (Taguchi, 1986 p. 83).) \square

The same design as given above can also be used for fewer than 7 factors, say 5, by simply omitting factors F and G , say. This allows us to estimate the experimental error by using the contrasts that would otherwise have been the estimates of the main effects F and G .

This last remark points out a potential difficulty with what are called *saturated main effect plans*, for example, a main effect plan for a 2^3 factorial in 4 runs or for a 2^7 factorial in 8 runs, do not allow estimation of error. Such information must then be obtained from external sources or the experiment must be enlarged by replication of at least some treatment combinations.

In the same way as described above we can examine 11 or fewer factors with 12 runs in a plan originated by Plackett and Burman (1946) or 15 or fewer factors with 16 runs using an orthogonal array, and so on. Construction of such designs will be discussed in Chapter II.14.

11.9 EXPERIMENTS WITH FACTORS AT THREE LEVELS

We have mentioned earlier the usefulness of 2^n factorial experiments, especially for exploratory studies. We have, however, pointed out also that 2^n factorials allow us to study relatively simple effect structures only. To be more specific, in studying quantitative factors we can make inferences only about linear (main) effects and linear \times linear-type interactions. This can easily be understood by writing the linear model as a

regression model. For example, for a 2^3 factorial in a CRD we write

$$\begin{aligned} y(x_1x_2x_3)_l = & \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 \\ & + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \beta_{123}x_1x_2x_3 \\ & + e(x_1x_2x_3)_l \quad (l = 1, 2, \dots, r), \end{aligned} \quad (11.46)$$

where x_1, x_2, x_3 represent the (coded) levels of factors A, B, C , respectively, with $x_i = -1, 1 (i = 1, 2, 3)$. In this model the regression coefficients, β_1, β_2 , and β_3 , are then (apart from a constant) the main effects of A, B, C , respectively, β_{12}, β_{13} , and β_{23} are the two-factor interactions between A and B , A and C , and B and C , respectively, and β_{123} is the 3-factor interaction between A, B, C . These seven regression coefficients account for the seven d.f. among the eight treatment combinations. There is, therefore, no opportunity to explore possible curvature in the main effects. For this we need at least three levels for each factor so that we have at least two d.f. for each main effect.

11.9.1 The 3^2 Factorial

Suppose then we have n factors A, B, C, \dots each at three levels. This is referred to as a 3^n factorial which can be used in conjunction with any error control design. Let us consider specifically the case $n = 2$ for purposes of illustration. Model (11.1) then reduces to

$$\tau_{ij} = \bar{\tau}_{..} + (\bar{\tau}_{i.} - \bar{\tau}_{..}) + (\bar{\tau}_{.j} - \bar{\tau}_{..}) + (\tau_{ij} - \bar{\tau}_{i.} - \bar{\tau}_{.j} + \bar{\tau}_{..})$$

or

$$\tau_{ij} = \mu + A_i + B_j + (AB)_{ij} \quad (11.47)$$

with $i, j = 0, 1, 2$ representing the levels of the factors A and B . The main effects for A and B account for two d.f. each and the interaction between factors A and B accounts for four d.f. making up the eight d.f. among the nine treatment combinations. These d.f. can be partitioned further (see Section 7.2) depending on the nature (that is, qualitative or quantitative) of the factors. We shall consider here the case where both factors are quantitative.

Let X_{1l} and X_{2l} ($l = 0, 1, 2$) denote the equally spaced levels of factors A and B , respectively. Then

$$x_{il} = \frac{X_{il} - \bar{X}_i}{\frac{1}{2}(X_{i2} - X_{i0})}$$

are the coded levels, with $x_{i0} = -1, x_{i1} = 0, x_{i2} = +1 (i = 1, 2)$. Similar to (11.46) we can then write a model of the form

$$\begin{aligned} y(x_{1l}x_{2l'})_m = & \beta_0 + \beta_1x_{1l} + \beta_2x_{2l'} + \beta_{11}x_{1l}^2 + \beta_{22}x_{2l'}^2 + \beta_{12}x_{1l}x_{2l'} \\ & + \beta_{122}x_{1l}x_{2l'}^2 + \beta_{112}x_{1l}^2x_{2l'} + \beta_{1122}x_{1l}^2x_{2l'}^2 + e(x_{1l}x_{2l'})_m \end{aligned} \quad (11.48)$$

Table 11.15 Design-Model Matrix \mathbf{X} for 3^2 Factorial

(x_1, x_2)	α_{00}	α_{10}	α_{01}	α_{20}	α_{02}	α_{11}	α_{12}	α_{21}	α_{22}
$(-1, -1)$	1	-1	-1	1	1	1	-1	-1	1
$(-1, 0)$	1	-1	0	1	-2	0	2	0	-2
$(-1, 1)$	1	-1	1	1	1	-1	-1	1	1
$(0, -1)$	1	0	-1	-2	1	0	0	2	-2
$(0, 0)$	1	0	0	-2	-2	0	0	0	4
$(0, 1)$	1	0	1	-2	1	0	0	-2	-2
$(1, -1)$	1	1	-1	1	1	-1	1	-1	1
$(1, 0)$	1	1	0	1	-2	0	-2	0	-2
$(1, 1)$	1	1	1	1	1	1	1	1	1

$(l, l' = 0, 1, 2; m = 1, 2, \dots, r)$. This is an explicit model accounting for all d.f. for main effects and two-factor interactions. Using the method of least squares, estimates of the regression coefficients can be obtained. Tests of hypotheses can be performed concerning these regression coefficients (see Chapter 4). Although this is straightforward, the interpretation is not always easy since the estimators are correlated. A more convenient way sometimes is to use a representation in terms of orthogonal polynomials (see Section 7.2).

Let $P_0(x)$, $P_1(x)$, and $P_2(x)$ be the zero-th, first, and second order polynomials, respectively, for $t = 3$ (see Table 7.3). We can then rewrite (11.48) as

$$\begin{aligned}
 y(x_{1l}, x_{2l'})_m &= \sum_{i,i'=0}^2 \alpha_{ii'} P_i(x_{1l}) P_{i'}(x_{2l'}) + e(x_{1l}, x_{2l'})_m \\
 &= \alpha_{00} + \alpha_{10} P_1(x_{1l}) + \alpha_{01} P_1(x_{2l'}) + \alpha_{20} P_2(x_{1l}) \\
 &\quad + \alpha_{02} P_2(x_{2l'}) + \alpha_{11} P_1(x_{1l}) P_1(x_{2l'}) \\
 &\quad + \alpha_{12} P_1(x_{1l}) P_2(x_{2l'}) + \alpha_{21} P_2(x_{1l}) P_1(x_{2l'}) \\
 &\quad + \alpha_{22} P_2(x_{1l}) P_2(x_{2l'}) + e(x_{1l} x_{2l'})_m. \tag{11.49}
 \end{aligned}$$

To make this representation more explicit it is useful to write (11.49) in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e}, \tag{11.50}$$

where \mathbf{y} is the column vector of observations, \mathbf{X} is the design-model matrix of known constants, $\boldsymbol{\alpha}$ is the column vector of regression coefficients, and \mathbf{e} is the column vector of errors. To simplify the notation we consider the case $r = 1$. The matrix \mathbf{X} is then as given in Table 11.15 (for $r > 1$, each row of \mathbf{X} is repeated r times). Since the columns of \mathbf{X} are orthogonal to each other, $\mathbf{X}'\mathbf{X}$ is a diagonal matrix, and it is then easy to obtain

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Table 11.16 ANOVA for Model (11.49)

Source	d.f.	SS	$E(\text{MS})$
Treatments	8	$\text{SS}(T)$	
A_L	1	$6r(\hat{\alpha}_{10})^2$	$\sigma_e^2 + 6r(\alpha_{10})^2$
B_L	1	$6r(\hat{\alpha}_{01})^2$	$\sigma_e^2 + 6r(\alpha_{01})^2$
A_Q	1	$18r(\hat{\alpha}_{20})^2$	$\sigma_e^2 + 18r(\alpha_{20})^2$
B_Q	1	$18r(\hat{\alpha}_{02})^2$	$\sigma_e^2 + 18r(\alpha_{02})^2$
$A_L \times B_L$	1	$4r(\hat{\alpha}_{11})^2$	$\sigma_e^2 + 4r(\alpha_{11})^2$
$A_L \times B_Q$	1	$12r(\hat{\alpha}_{12})^2$	$\sigma_e^2 + 12r(\alpha_{12})^2$
$A_Q \times B_L$	1	$12r(\hat{\alpha}_{21})^2$	$\sigma_e^2 + 12r(\alpha_{21})^2$
$A_Q \times B_Q$	1	$36r(\hat{\alpha}_{22})^2$	$\sigma_e^2 + 36r(\alpha_{22})^2$
Error	$9(r-1)$	$\text{SS}(E)$	σ_e^2
Total	$9r-1$	$\text{SS}(\text{Total})$	

For example,

$$\hat{\alpha}_{10} = \frac{1}{6}[Y(1, \cdot) - Y(-1, \cdot)]$$

where $Y(1, \cdot) = \sum_{l'=0}^2 y(1, x_{2l'})$, and so on. Obviously, $\hat{\alpha}_{10}$ is the estimator for the linear effect of factor A , A_L say. Similarly,

$$\hat{\alpha}_{20} = \frac{1}{18}[Y(-1, \cdot) - 2Y(0, \cdot) + Y(1, \cdot)]$$

is the estimator for the quadratic effect (that is, curvature) of factor A , A_Q say. To mention one of the interaction parameters, consider

$$\hat{\alpha}_{11} = \frac{1}{4}\{[y(1, 1) - y(1, -1)] - [y(-1, 1) + y(-1, -1)]\},$$

that is, a comparison of the linear effect of B at $x_1 = 1$ versus the linear effect of B at $x_1 = -1$. We denote this interaction by $A_L \times B_L$. Other interaction effects are defined and estimated similarly.

Test of hypotheses about the regression coefficients in (11.49) can be made in conjunction with the ANOVA by partitioning $\text{SS}(T)$ into eight single d.f. sums of squares each accounting for one of the regression coefficients. Details are given in Table 11.16. We note that the sums of squares are given for a CRD with $r > 1$ replications, with the appropriate changes in the estimated regression coefficients, for example,

$$\hat{\alpha}_{10} = \frac{1}{6r}[Y(1, \cdot) - Y(-1, \cdot)]$$

where

$$Y(1, \cdot) = \sum_{l'=0}^2 \sum_{m=1}^r y(1, x_{2l'})_m$$

is the sum of all the observations with $x_1 = 1$, etc. The method of computing the sums of squares is, of course, that given in (7.24).

11.9.2 Extensions

Extensions of this method to the general case of n factors each at three levels should now be obvious. As more factors are included in the experiment many more d.f. become available for interactions, not only for interactions between two factors, but three factors, four factors, and so on. For higher order interactions partitions into $A_L \times B_L \times C_L$, $A_L \times B_L \times C_Q$, and so forth may not be particularly useful as these components will become difficult to interpret except that they are part of the $A \times B \times C$ interaction. Moreover, just as with 2^n factorials, it is entirely likely that interactions involving three or more factors are negligible and that their sums of squares may be pooled with $SS(E)$.

11.9.3 Formal Definition of Main Effects and Interactions

Our discussion of 3^n factorials so far has concentrated on quantitative factors with equally spaced levels. A more general representation dealing with other situations and in particular qualitative factors is obviously needed. We shall discuss such a method here briefly, deferring a more in-depth discussion to Chapter II.10.

We have seen in Section 11.3 that for the 2^n factorial each main effect and each interaction can be expressed as a contrast among all 2^n treatment combinations or, more precisely, among the true responses of all 2^n treatment combinations. Thus each main effect and interaction is represented by a single d.f. contrast. Moreover, these contrasts are mutually orthogonal.

For the 3^n factorial we have seen above that each main effect, A say, consists of two comparisons, A_L and A_Q . And each 2-factor interaction, $A \times B$ say, consists of four contrasts, $A_L \times B_L$, $A_L \times B_Q$, $A_Q \times B_L$ and $A_Q \times B_Q$. Extending this, each 3-factor interactions consist of 8 comparisons, and so on. Expressed alternatively, each main effect is represented by or accounts for 2 d.f., each 2-factor interaction for 4 d.f., each 3-factor interaction for 8 d.f., and so on. To partition the 4 d.f. for 2-factor interactions into two orthogonal sets of 2 d.f. each and the 8 d.f. for 3-factor interactions into 4 mutually orthogonal sets of 2 d.f. each, Yates (1937) introduced what we might call *interaction orthogonal components*. This notion and the mathematics of the method were formalized by Kempthorne (1952) and can be described as follows.

EXAMPLE 11.8: Let us consider the 3^3 factorial. We write a treatment combination as $\mathbf{x} = (x_1, x_2, x_3)$ where x_i denotes the level of the i th factor with $x_i = 0, 1, 2$ ($i = 1, 2, 3$). To simplify the notation we denote the three factors by A , B , and C . Further, let $\tau(\mathbf{x})$ represent the true effect of the treatment combination \mathbf{x} .

We know from Section 11.9.1 that the 2 d.f. corresponding to main effect A , say, are represented by comparisons among the treatment means with $x_1 = 0, x_1 = 1$, and $x_1 = 2$, that is, $\bar{\tau}(0, \cdot, \cdot)$ vs. $\bar{\tau}(1, \cdot, \cdot)$ vs. $\bar{\tau}(2, \cdot, \cdot)$. Here $\bar{\tau}(x_1, \cdot, \cdot)$ for $x_1 = 0, 1, 2$ is a mean of 9 effects averaging over the 9 treatment combinations (x_1, x_2, x_3) with $x_2, x_3 = 0, 1, 2$ and fixed x_1 . Similarly, the main effects for B and C are represented by comparisons among treatment means with $x_2 = 0, x_2 = 1, x_2 = 2$, and $x_3 = 0, x_3 = 1, x_3 = 2$, respectively.

This idea of two comparisons among three treatment means or, alternatively, among three sets of treatment combinations can be carried over to the various interactions. As mentioned above the 2-factor interaction $A \times B$ is partitioned into two components. These components are denoted by AB and AB^2 and are defined as follows: AB is represented by comparisons among the three means of treatments satisfying the equations

$$x_1 + x_2 = 0 \text{ vs. } x_1 + x_2 = 1 \text{ vs. } x_1 + x_2 = 2,$$

where $x_1, x_2 = 0, 1, 2$ and all arithmetic is modulo 3. The second component, AB^2 , is represented by comparisons among means of treatments satisfying the equations

$$x_1 + 2x_2 = 0 \text{ vs. } x_1 + 2x_2 = 1 \text{ vs. } x_1 + 2x_2 = 2$$

all mod 3. The remaining 2-factor interaction components, AC and AC^2 for $A \times C$, and BC and BC^2 for $B \times C$, are defined similarly. Finally, the four interaction components of the 3-factor interaction $A \times B \times C$ are denoted by $ABC, AB^2C, ABC^2, AB^2C^2$, and are represented by comparisons among sets of treatment combinations satisfying the following equations

$$\begin{aligned} ABC : \quad x_1 + x_2 + x_3 &= 0, 1, 2, \text{ mod } 3 \\ AB^2C : \quad x_1 + 2x_2 + x_3 &= 0, 1, 2, \text{ mod } 3 \\ ABC^2 : \quad x_1 + x_2 + 2x_3 &= 0, 1, 2, \text{ mod } 3 \\ AB^2C^2 : \quad x_1 + 2x_2 + 2x_3 &= 0, 1, 2, \text{ mod } 3. \end{aligned}$$

We summarize the above definition of the various main effects and interactions for the 3^3 factorial in Table 11.17, giving the names of the effects/interactions, and interaction components together with their d.f. and the left-hand sides of the equations defining the partitions of the 3^3 treatment combinations (the right-hand sides are always 0, 1, 2 mod 3). The reader should have no difficulties extending the procedure to 3^n factorials with $n > 3$.

To motivate these definitions of main effects and interaction components we referred to contrasts defining linear and quadratic contrasts as parts of main effects, for example, A_L and A_Q as parts of the main effect for factor A . In the context of the above discussion A_L is represented by the comparison of treatment combinations satisfying $x_1 = 0$ vs. $x_1 = 2$. Similarly, A_Q is represented by the comparison of the form $\frac{1}{2}\{x_1 = 0 \text{ and } x_1 = 2\}$ vs. $x_1 = 1$. We know, of course, that the two contrasts are orthogonal. For the formal definition as given above this does not have to be the case. For example, the two comparisons for main effect A could be $\{x_1 = 0 \text{ vs. } x_1 = 1\}$ and $\{x_1 = 0 \text{ vs. } x_1 = 2\}$. Another point we need to make here is that the comparisons for, say, AB and AB^2 bear no relationship to $A_L \times B_L, A_L \times B_Q, A_Q \times B_L$, and $A_Q \times B_Q$.

Table 11.17 Main Effects and Interactions for the 3^3 Factorial

Main Effect/Interaction	d.f.	Equation
A	2	x_1
B	2	x_2
$A \times B$	4	
AB	2	$x_1 + x_2$
AB^2	2	$x_1 + 2x_2$
C	2	x_3
$A \times C$	4	
AC	2	$x_1 + x_3$
AC^2	2	$x_1 + 2x_3$
$B \times C$	4	
BC	2	$x_2 + x_3$
BC^2	2	$x_2 + 2x_3$
$A \times B \times C$	8	
ABC	2	$x_1 + x_2 + x_3$
AB^2C	2	$x_1 + 2x_2 + x_3$
ABC^2	2	$x_1 + x_2 + 2x_3$
AB^2C^2	2	$x_1 + 2x_2 + 2x_3$

These two representations simply refer to different partitions of the 4 d.f. for $A \times B$. □

The formal definitions of main effects and interaction components as presented in this section are most valuable in considering suitable arrangements for 3^n factorials in incomplete blocks or in choosing suitable fractions of 3^n factorials. We shall illustrate this with a few simple examples.

11.9.4 Systems of Confounding for the 3^n Factorial

Let us consider the 3^3 factorial and blocks of size 9 ($= 3^2$). To accommodate all 27 treatment combinations in blocks of size 9 we need 3 blocks. The idea then is, just as in the case of 2^n factorial (see Section 11.6), to confound certain interactions with blocks. In our case we have 2 d.f. among 3 blocks, which means that we need to confound an interaction with 2 d.f. with blocks. More precisely, we shall choose an interaction component which does account for 2 d.f. as we have just explained. If, for example, the 3-factor interaction $A \times B \times C$ is assumed to be negligible or unimportant we can

Table 11.18 Partition of 3^3 Treatment Combinations

Set 1: $x_1 + x_2 + x_3 = 0$	Set 2: $x_1 + x_2 + x_3 = 1$	Set 3: $x_1 + x_2 + x_3 = 2$
0 0 0	1 0 0	2 0 0
1 1 1	2 1 1	0 1 1
2 2 2	0 2 2	1 2 2
1 2 0	2 2 0	0 2 0
2 1 0	0 1 0	1 1 0
1 0 2	2 0 2	0 0 2
2 0 1	0 0 1	1 0 1
0 1 2	1 1 2	2 1 2
0 2 1	1 2 1	2 2 1

choose any one of the four interaction components ABC , AB^2C , ABC^2 , or AB^2C^2 to confound with blocks. They are all equally important or, in our case, unimportant. Suppose we choose ABC . We then partition the 3^3 treatment combinations into three sets according to the equations for ABC , namely:

- Set 1: $x_1 + x_2 + x_3 = 0 \pmod 3$
- Set 2: $x_1 + x_2 + x_3 = 1 \pmod 3$
- Set 3: $x_1 + x_2 + x_3 = 2 \pmod 3$.

Each set contains, of course, exactly 9 treatment combinations. These sets are given in Table 11.18. The treatment combinations in a given set are then all assigned to the same block thus generating what we shall refer to as the basic arrangement in 3 blocks of size 9 each. For the actual experiment the basic arrangement may then be replicated r times giving us $3r$ blocks.

Using the formal definitions of Section 11.9.3 we can verify easily that only ABC is confounded with blocks and that all other main effects and interactions can be estimated from this arrangement. For example, in each set (block) there are exactly three treatment combinations with $x_1 = 0$, $x_1 = 1$, and $x_1 = 2$ which implies that any comparison among these three sets of treatment combinations is free of block effects.

A consequence of this system of confounding is that although full information is obtained on all main effects and 2-factor interactions, only limited information on the 3-factor interaction $A \times B \times C$ is available through the interaction components AB^2C , ABC^2 , and AB^2C^2 . This means that in the ANOVA table the 3-factor interaction $A \times B \times C$ has only 6 d.f.

This may be sufficient to get some idea whether 3-factor interaction is present. If, on the other hand, one is not willing to make the assumption that there is no interaction (as we did above) an obvious solution would be to use partial confounding (see Section 11.6.3). This can be done in various ways. We shall mention only two to illustrate

the general idea:

- (i) Use two basic replicates by confounding two of the 3-factor interaction components, say ABC and AB^2C , one in each of the basic replicates. This allows $1/2$ information on ABC and AB^2C and full information on ABC^2 and AB^2C^2 , thus restoring the 8 d.f. for $A \times B \times C$.
- (ii) If sufficient experimental material is available we could use four basic replicates confounding ABC , AB^2C , ABC^2 , and AB^2C^2 in one of the four basic replicates, respectively. This will yield then $3/4$ information on all four components and hence on the 8 d.f. for $A \times B \times C$.

The general ideas presented in Section 11.6.2 to construct systems of confounding for 2^n factorials in blocks of size 2^l ($l < n$) can be extended to 3^n factorials in blocks of size 3^l . For example, we may consider a 3^3 factorial in blocks of size 3. Obviously, several main effects or interaction components will have to be confounded with blocks; in this case four to be exact since we have 9 blocks and hence 8 d.f. among the blocks. These have to be chosen carefully. We shall not pursue this here any further but defer description of the general procedure to Chapter II.10.

11.9.5 Fractions of 3^n Factorials

Even more so than for the 2^n factorial the number of treatment combinations for the 3^n factorial may be far too large for practical applications. And again, if higher order interactions are considered to be negligible it may be entirely satisfactory to consider only a fraction of all possible treatment combinations and still obtain most if not all of the information needed. The easiest method is to consider $1/3^l$ fractions of the 3^n factorial ($l < n$). We shall give a simple example here and defer a more general description to Chapters II.13 and 14.

EXAMPLE 11.9: Let us consider a $1/3$ fraction of the 3^3 factorial, that is, 9 out of the possible 27 treatment combinations. As in Section 11.7 the general idea is to assume that the highest-order interaction, in this case the 3-factor interaction, is negligible and use that fact to choose the treatment combinations to be included in the fraction. Recall that for each interaction component the set of all treatment combinations is partitioned into three subsets. For example for the component ABC the subsets will be obtained by satisfying the equations

$$x_1 + x_2 + x_3 = 0, 1, 2 \text{ mod } 3.$$

Any one of these subsets constitutes a $1/3$ fraction which obviously does not allow estimation of contrasts belonging to ABC . What are other consequences?

Since we are considering only 9 treatment combinations we have 8 d.f. for treatments, that is, there are 8 linearly independent comparisons among treatment effects that can be estimated. How can we identify these? Let us consider set 1 in Table 11.18

which constitutes the $1/3$ fraction under consideration here:

$$\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 2 & 0 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 2 & 1 & 0 & 2 & 0 & 2 & 1 \end{array}$$

The 2 d.f. associated with the main effect for factor A result from the comparisons of treatment combinations satisfying

$$\{x_1 = 0\} \quad \text{vs.} \quad \{x_1 = 1\} \quad \text{vs.} \quad \{x_1 = 2\}, \quad (11.51)$$

that is,

$$\begin{Bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{Bmatrix} \quad \text{vs.} \quad \begin{Bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{Bmatrix} \quad \text{vs.} \quad \begin{Bmatrix} 2 & 2 & 2 \\ 2 & 1 & 0 \\ 2 & 0 & 1 \end{Bmatrix}. \quad (11.52)$$

A close look at the three sets of treatment combinations given in (11.52) shows that in addition to satisfying the equations in (11.51) they also represent comparisons satisfying the equations

$$\{x_2 + x_3 = 0\} \quad \text{vs.} \quad \{x_2 + x_3 = 2\} \quad \text{vs.} \quad \{x_2 + x_3 = 1\} \quad (11.53)$$

and

$$\begin{aligned} \{x_1 + 2x_2 + 2x_3 = 0\} \quad \text{vs.} \quad \{x_1 + 2x_2 + 2x_3 = 2\} \\ \text{vs.} \quad \{x_1 + 2x_2 + 2x_3 = 1\}. \end{aligned} \quad (11.54)$$

This means that the contrasts belonging to A also belong to the interaction components BC and AB^2C^2 as indicated by (11.53) and (11.54), respectively (see also Table 11.17). In the terminology of Section 11.7 we thus say that A , BC and AB^2C^2 are confounded or aliased with each other.

Using similar arguments we can also show that B , AC , and AB^2C are confounded with each other, and so are C , AB , ABC^2 and, finally, AB^2 , AC^2 , BC^2 . To summarize, the alias structure for this fraction can be written as (using the convention of Section 11.7):

$$\begin{aligned} A &= BC = AB^2C^2 \\ B &= AC = AB^2C \\ C &= AB = ABC^2 \\ AB^2 &= AC^2 = BC^2. \end{aligned} \quad (11.55)$$

We have thus identified four sets of comparisons, each accounting for 2 d.f. In addition these four sets are linearly independent of each other which means that we have indeed identified the 8 comparisons accounting for the 8 d.f. among the 9 treatment combinations.

Formally, (11.55) can be derived by a mathematical argument similar to that described in Section 11.7. We start with the identity relationship which determines the fraction, that is, the treatment combinations to be chosen as well as the alias structure. For our example the identity relationship is

$$I = ABC \quad (11.56)$$

or if we want to determine the fraction uniquely

$$I = ABC_0,$$

which means that we choose the treatment combination satisfying $x_1 + x_2 + x_3 = 0$ rather than the other two possibilities, that is, $x_1 + x_2 + x_3 = 1$ or $x_1 + x_2 + x_3 = 2$. To obtain the alias structure we proceed as follows:

- (i) We consider (11.56) as a mathematical equation with I being the identity.
- (ii) We multiply each effect, that is, main effect or interaction component, formally into both sides of (11.56);
- (iii) The power for each letter is reduced mod 3 if necessary and any letter raised to power 0 is deleted from the expression.
- (iv) If the first letter with nonzero power is raised to the power 2 then the entire expression is squared and again reduced mod 3 (this is done to adhere to the convention for having a unique enumeration of all possible effects such as given in Table 11.17).
- (v) In addition each effect is multiplied in the same way into $(ABC)^2$ to obtain the second alias.

To illustrate these steps we shall find the aliases of A , namely,

$$\begin{aligned}
 A &= A(ABC) = A(ABC)^2 \\
 &= A^2BC = A^3B^2C^2 \\
 &= (A^2BC)^2 = B^2C^2 \\
 &= A^4B^2C^2 = (B^2C^2)^2 \\
 &= AB^2C^2 = B^4C^4 \\
 &= AB^2C^2 = BC.
 \end{aligned} \quad (11.57)$$

In Section 11.6.2 we have referred to the generalized interaction, XY say, for two effects X and Y in the 2^n system. In the 3^n system we have for any two effects X and Y , say, not only one but two generalized interactions, denoted by XY and XY^2 . Using the concept of the generalized interaction we can then also say that, for example, A is aliased with the generalized interactions of A and ABC , as given in (11.57). The

remaining aliases in (11.56) can be obtained in the same way. Finally, we note that together with ABC in (11.56) all possible effects for the 3^3 factorial (see Table 11.17) are accounted for in (11.55). \square

The fractional factorial design we have discussed above is clearly of no practical value unless in addition to the assumption of zero 3-factor interaction we can also assume that all 2-factor interactions are zero or negligible. Then we can obtain information about main effects. This is another example of a main-effect plan or a resolution III design.

11.9.6 Highly Fractionated 3^n Factorials

The reader should have no difficulty extending the ideas of the previous section and consider, for example, a $1/3$ fraction of the 3^4 factorial leading to a resolution IV design or a $1/3$ fraction of the 3^5 factorial leading to a resolution V design, and so on. But even a $1/3$ fraction of a 3^5 factorial may be impractical as it leads still to too many treatment combinations. The problem becomes even more critical for 3^n factorials with larger n . And it is not uncommon to have many factors in an experiment, in particular in an exploratory experiment. The need for highly fractionated factorials becomes then obvious, such as a $1/9$ fraction of a 3^5 or 3^6 , or a $1/27$ fraction of a 3^6 factorial, and so on.

Designs of the form mentioned above can be developed by combining and extending the ideas and rules given in Sections 11.7.4 and 11.9.5. In particular, the identity relationship now contains several interaction components; some are chosen independently and others represent the generalized interactions of those chosen interaction components. For example, for a $1/3^2$ fraction we can choose two interactions freely, say X and Y , so that

$$I = X = Y = XY = XY^2$$

determines the treatment combinations to be included and also the alias structure. It is, of course, important to have an alias structure which allows us, under certain assumptions, to estimate the effects and interactions in which we are interested. This is not always easy to do and care must be used to choose X and Y appropriately, or X, Y, Z , say, for a $1/3^3$ fraction, etc. We shall not pursue this any further here, but some rules will be developed in Chapters II.13 and 14.

11.9.7 Systems of Confounding for Fractions of 3^n Factorials

Even for a reasonable fractional factorial the number of treatment combinations may be too large for a suitable error-control design. For example, for the $1/3$ fraction of the 3^3 factorial we have 9 treatment combinations but the error-control design available may call for blocks of size 3. It becomes then necessary to use an incomplete block design and confound some effects with blocks, that is, use a system of confounding as described in Section 11.9.4.

To obtain a reasonable, that is, useful system of confounding we have to choose carefully the effect or effects to be confounded with blocks in order not to sacrifice needed information. To make that choice we have to consult the alias structure to see which effects can be estimated (if there were no blocking) and which of these are least important. These would be typically higher order interactions. Let us consider our example of Section 11.9.5 and suppose we have available blocks of size 3. We then need three blocks for a basic replicate. That means we need to confound 2 d.f. with blocks and since each effect in a 3^n factorial accounts for 2 d.f. we need to confound one effect with blocks. Inspection of the alias structure (11.55) shows that if we do not want to confound a main effect with blocks the only choice is to confound AB^2 (and its aliases) with blocks. Using the procedure of Section 11.9.4 we construct the blocks by finding the treatment combinations (among the 9 chosen for the fraction) satisfying the equations

$$x_1 + 2x_2 = 0 : 000, 111, 222$$

$$x_1 + 2x_2 = 1 : 210, 102, 021$$

$$x_1 + 2x_2 = 2 : 120, 201, 012$$

and assign them to blocks 1, 2, 3, respectively. Suppose we have r replications of this basic arrangement, that is, $3r$ blocks altogether. Then the structure of the ANOVA is as given in Table 11.19. The important point is that there are now only 6 d.f. for treatments which are partitioned into the main effects A , B , and C each with 2 d.f.

This simple example should convey the general idea that the construction of systems of confounding for fractional factorials follows the same rules as for full factorials. The effects to be confounded are obtained from the alias structure. With each effect its aliases are also confounded with blocks. And if several effects need to be confounded with blocks then all their generalized interactions are confounded with blocks too. This makes this process not always easy and as a consequence sometimes confounding of desirable effects cannot be avoided. Systems of partial confounding may be helpful.

11.10 EXPERIMENTS WITH FACTORS AT TWO AND THREE LEVELS

11.10.1 Asymmetrical Factorial Experiments

So far we have discussed two extreme types of factorial experiments: On the one extreme we have n factors all with possibly different numbers of levels; on the other extreme we have n factors all having the same number of levels, for instance, 2 or 3. These two types are referred to as *asymmetrical (mixed) factorials* and *symmetrical (pure) factorials*, respectively.

In practical applications special kinds of asymmetrical factorials are often used. We may have two or three groups of factors where all factors in the same group have the same number of levels. Of particular interest are $2^m \times 3^n$ experiments, that is, m factors

Table 11.19 Outline of ANOVA for 1/3 Fraction of 3^3 Factorial in Blocks of Size 3

Source	d.f.
Blocks	$3r - 1$
<i>A</i>	2
<i>B</i>	2
<i>C</i>	2
Error	$6(r - 1)$
Total	$9r - 1$

with 2 levels each and n factors with 3 levels each. Their use has been advocated and promoted by Taguchi (1986, 1987) (see also Roy, 1990) in his parameter designs for off-line quality control. Of special importance in these applications are fractions of $2^m \times 3^n$ factorials (see Section II.17.4.1).

The construction of such fractions and of systems of confounding borrows heavily from the methods we have discussed in earlier sections. We shall give a few examples to illustrate the main ideas, but leave a more thorough discussion for Chapters II.12 and 13.

11.10.2 Confounding in $2^m \times 3^n$ Factorials

To use the methods of constructing systems of confounding described in Sections 11.6 and 11.9.4 we need to confine ourselves to blocks of size $2^p \times 3^q$ with $p \leq m$, $q \leq n$. The general idea is to either combine a system of confounding for the 2^m factorial with the complete 3^n factorial, or a system of confounding for the 3^n factorial with the complete 2^m factorial or, as a third possibility, combine systems of confounding for both factorials. We shall illustrate this for the $2^2 \times 3^2$ factorial with blocks of size 18, 12, 9, 6, and 4.

Let us denote the treatment combinations by (x_1, x_2, z_1, z_2) where $x_1, x_2 = 0, 1$ represent the levels of the 2^2 factorial with factors A, B and $z_1, z_2 = 0, 1, 2$ those of the 3^2 factorial with factors C, D . Further, let S_i denote the i th set of treatment combinations for a system of confounding for the 2^2 factorial and S'_j the j th set of a system of confounding for the 3^2 factorial. Combining sets S_i and S'_j in an appropriate way, referred to as a *Kronecker product design*, constitutes then a system of confounding for the $2^2 \times 3^2$ factorial. These can be described briefly as follows.

Blocks of Size 18:

Confounding AB with blocks of size 2 gives $S_1 = \{(0, 0), (1, 1)\}$, $S_2 = \{(1, 0), (0, 1)\}$. With $S' = \{(0, 0), (1, 0), (2, 0), (0, 1), (1, 1), (2, 1), (0, 2), (1, 2), (2, 2)\}$ we then consider S_1S' and S_2S' . This means we adjoin every treatment combination in S_i ($i = 1, 2$) with every treatment combination in S' , giving us two sets of 18 treatment combinations (x_1, x_2, z_1, z_2) . Each set represents a block. These two blocks form the basic arrangement which can then be replicated r times. Except for the interaction AB (1 d.f.) the main effects for A, B, C, D and all other interactions are estimable.

An alternative to replicating the basic arrangement is to use partial confounding of A, B , and AB . Such a plan yields then partial information about these three effects and full information about all other effects.

Blocks of Size 12:

For this situation we generate three sets S'_1, S'_2, S'_3 by confounding, for example, CD . We then have $S'_1 = \{(0, 0), (1, 2), (2, 1)\}$, $S'_2 = \{(1, 0), (0, 1), (2, 2)\}$, and $S'_3 = \{(2, 0), (0, 2), (1, 1)\}$. With $S = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ we form SS'_1, SS'_2, SS'_3 which yields three blocks of size 12. This basic arrangement needs to be replicated r times. Alternatively, some system of partial confounding for the 3^2 factorial may be used so that information about all main effects and interactions may be obtained.

Blocks of Size 9:

The only design in this class is obtained by confounding A, B , and AB , that is, by forming $S_1 = \{(0, 0)\}$, $S_2 = \{(1, 0)\}$, $S_3 = \{(0, 1)\}$, $S_4 = \{(1, 1)\}$. These sets are then combined with $S' = \{\text{all treatment combinations for } 3^2 \text{ factorial}\}$. This arrangement is obviously of no practical value unless the 2^2 factorial itself is not important but only the 3^2 factorial and interactions between factors with 2 and 3 levels, for example, $A \times C, B \times C, A \times B \times C$, etc.

Blocks of Size 6:

This is the only situation where we combine systems of confounding for both the 2^2 and 3^2 factorials. One possibility is to confound AB generating $S_1 = \{(0, 0), (1, 1)\}$ and $S_2 = \{(1, 0), (0, 1)\}$, and to confound CD , say, generating $S'_1 = \{(0, 0), (1, 2), (2, 1)\}$, $S'_2 = \{(1, 0), (0, 1), (2, 2)\}$ and $S'_3 = \{(2, 0), (0, 2), (1, 1)\}$. The six combinations $S_iS'_j$ ($i = 1, 2; j = 1, 2, 3$) then yield six blocks of size 6. We should note here that in addition to AB and CD also the generalized interaction $AB \times CD$ is confounded with blocks. There exist, obviously, other possibilities of forming the S_i and S'_j and various system of partial confounding can be used to obtain the desired amount of information about main effects and interactions.

Blocks of Size 4:

As with the case of blocks of size 9, this design is generally of no practical value as all effects of the 3^2 are confounded with blocks. We combine $S_1 = \{(0, 0), (1, 0)\}$,

$(0, 1), (1, 1)\}$ with S'_j ($j = 1, 2, \dots, 9$) where each S'_j contains only one treatment combination from the 3^2 factorial.

The method just described is obviously quite simple and can be extended easily to other factorials. It does, however, not always lead to the most practical or suitable arrangements. Using a different notion of partial confounding and employing other types of incomplete block designs we shall discuss other methods in Chapter II.12 (for a listing of some useful designs see II. Appendix D).

11.10.3 Fractions of $2^m \times 3^n$ Factorials

The idea of considering the symmetrical factorials separately and then adjoining treatment combinations from those factorials in an appropriate manner can also be used to construct useful fractions of asymmetrical factorials. Connor and Young (1961) have devised such a method for $2^m \times 3^n$ factorials. Their designs are such that they allow the estimation of all main effects and 2-factor interactions assuming that all other interactions are negligible.

We shall give only one example here to illustrate the method and refer the reader to the catalog of designs provided by Connor and Young (1961) as reprinted in McLean and Anderson (1984). We consider a $1/2$ fraction of the $2^3 \times 3^2$ factorial with factors A, B, C having 2 levels and D, E having 3 levels. To this end we consider a $1/2$ fraction of the 2^3 factorial based on the identity relationship

$$I = ABC.$$

This leads to a partition of the 8 treatment combinations into two sets, S_1 and S_2 , according to the sign with which the treatment combinations into ABC (see Section 11.6) or, alternatively, according to whether they satisfy the equations

$$S_1 : x_1 + x_2 + x_3 = 0 \bmod 2$$

or

$$S_2 : x_1 + x_2 + x_3 = 1 \bmod 2.$$

We thus obtain Each set represents a $1/2$ fraction of the 2^3 factorial.

S_1	S_2
0 0 0	1 0 0
1 1 0	0 1 0
1 0 1	0 0 1
0 1 1	1 1 1

For the 3^2 factorial we consider $1/3$ fractions based on the identity relationship

$$I = DE.$$

This leads to a partition into three sets S'_1, S'_2, S'_3 based on the equations

$$S'_1 : z_1 + z_2 = 0 \bmod 3$$

$$S'_2 : z_1 + z_2 = 1 \bmod 3$$

$$S'_3 : z_1 + z_2 = 2 \bmod 3,$$

that is,

S'_1	S'_2	S'_3
0 0	1 0	2 0
1 2	0 1	0 2
2 1	2 2	1 1

The $1/2$ fraction for the $2^3 \times 3^2$ factorial is then obtained by adjoining the two types of sets as follows:

$$S_1 S'_1, \quad S_2 S'_2, \quad S_2 S'_3$$

Each set $S_i S'_j$ consists of $4 \times 3 = 12$ treatment combinations. The final design is given in Table 11.20.

Since both $1/2$ fractions of the 2^3 factorial and all three $1/3$ fractions of the 3^2 factorial are used to obtain the final design it is possible to estimate all main effects and all 2-factor interactions. As a consequence these types of designs are still quite large and other designs may have to be considered for practical applications. Of particular interest then are main effect plans as developed, for example, by Addelman and Kempthorne (1961) and Addelman (1962). Such methods are discussed in Chapter II.14.

Table 11.20 $1/2$ Fraction of $2^3 \times 3^2$ Factorial

$S_1 S'_1$	$S_2 S'_2$	$S_2 S'_3$
00000	10010	10020
00012	10001	10002
00021	10022	10011
11000	01010	01020
11012	01001	01002
11021	01022	01011
10100	00110	00120
10112	00101	00102
10121	00122	00111
01100	11110	11120
01112	11101	11102
01121	11122	11111

11.11 EXAMPLES USING SAS®

EXAMPLE 11.10: We consider here a purely numerical example to illustrate the roles of error-control, treatment and sampling design. We use a design of partial confounding of a 2^2 factorial in blocks of size 2 with subsampling (two observations per EU). The design and the data are given in Table 11.21b.

We use both SAS PROC GLM and SAS PROC MIXED to analyze the data. The main reason for using PROC GLM is to obtain an ANOVA table. The results of both analyses are given in Table 11.21b, based on the input statements given in Table 11.21a.

We make the following comments on the input and output:

- (i) In order to obtain the correct test for A , B , $A * B$ we have to specify in PROC GLM $E = \text{block} * A * B$ as the correct error term, that is, the experimental error. In PROC MIXED this is done correctly automatically by declaring $\text{block} * A * B$ as a random effect.
- (ii) The observational error variance component is estimated as $\hat{\sigma}_\eta^2 = 1.2917$ (in GLM in the basic ANOVA and in MIXED as Residual).
- (iii) The experimental error variance component is estimated in MIXED as $\hat{\sigma}_\epsilon^2 = \text{block} * A * B = 1.7292$. We can obtain the same value in GLM from MS ($\text{block} * A * B$) as

$$\begin{aligned}\hat{\sigma}_\epsilon^2 &= [\text{MS}(\text{block} * A * B) - \text{MS}(\text{ERROR})]/2 \\ &= (4.75 - 1.2917)/2 = 1.7292\end{aligned}$$

- (iv) Both analyses produce the same results for testing hypotheses about A , B and $A * B$.
- (v) The estimates for A , B , and $A * B$ are obtained in MIXED by specifying the appropriate contrasts. Since all effects are confounded in one (out of three) replicates, they are all estimated with the same variance, namely,

$$\begin{aligned}\text{var}(\hat{A}) = \text{var}(\hat{B}) &= \text{var}(\widehat{A * B}) \\ &= \frac{\sigma_\eta^2 + n\sigma_\epsilon^2}{r^*n},\end{aligned}$$

where r^* is the effective number of replications and n is the size of the subsam-

ple. As a consequence,

$$\begin{aligned}
 se(\hat{A}) = se(\hat{B}) &= se(\widehat{A * B}) \\
 &= \left(\frac{MS(E E)}{r * n} \right)^{1/2} \\
 &= \left(\frac{MS(\text{block} * A * B)}{4} \right)^{1/2} \\
 &= \left(\frac{4.75}{4} \right)^{1/2} = 1.0897.
 \end{aligned}$$

(vi) The LS means (which are not available in GLM) can also be used to obtain the estimates for A , B , and $A * B$.

(vii) The d.f. for testing hypotheses about A , B , and $A * B$ are the d.f. for experimental error. They are obtained as

$$\# \text{ of EUs} - \# \text{ of treatments} - \# \text{ of blocks} + 1$$

$$= 12 - 4 - 6 + 1$$

$$= 3 \quad (\text{see Table 9.13}).$$

□

Table 11.21 2^2 Factorial in Blocks of Size 2

a) Input statements:

```

data factorial;
input block A B y @@;
datalines;
1 0 4 1 0 5 1 1 7 1 1 6
2 1 6 2 1 8 2 0 1 10 2 0 1 1
3 1 0 7 3 1 0 7 3 0 0 3 3 0 0 5
4 0 1 9 4 0 1 12 4 1 1 12 4 1 1 14
5 0 1 10 5 0 1 11 5 0 0 6 5 0 0 5
6 1 0 7 6 1 0 6 6 1 1 8 6 1 1 10
;
run;

proc print data=factorial;
title1 'DATA FOR 2**2 FACTORIAL';
title2 'IN INCOMPLETE BLOCKS';
title3 'WITH SUBSAMPLING';
run;

proc glm data=factorial;
class block A B;
model y = block A B A*B block*A*B;
test H = A B A*B E = block*A*B;
title1 'ANALYSIS OF 2**2 FACTORIAL';
run;

```

Table 11.21 (Continued)

```
proc mixed data=factorial;
class block A B;
model y = block A B A*B;
random block*A*B;
lsmeans A B A*B;
estimate 'A' A -1 1;
estimate 'B' B -1 1;
estimate 'A*B' A*B -1 1 1 -1/divisor=2;
run;
```

b.) Output:

DATA FOR 2**2 FACTORIAL IN INCOMPLETE BLOCKS WITH SUBSAMPLING				
Obs	block	A	B	y
1	1	0	0	4
2	1	0	0	5
3	1	1	1	7
4	1	1	1	6
5	2	1	0	6
6	2	1	0	8
7	2	0	1	10
8	2	0	1	11
9	3	1	0	7
10	3	1	0	7
11	3	0	0	3
12	3	0	0	5
13	4	0	1	9
14	4	0	1	12
15	4	1	1	12
16	4	1	1	14
17	5	0	1	10
18	5	0	1	11
19	5	0	0	6
20	5	0	0	5
21	6	1	0	7
22	6	1	0	6
23	6	1	1	8
24	6	1	1	10

ANALYSIS OF 2**2 FACTORIAL	
The GLM Procedure	
Class Level Information	
Class	Levels Values
block	6 1 2 3 4 5 6
A	2 0 1
B	2 0 1
Number of Observations Read	
Number of Observations Used	
24	
24	

Table 11.21 (Continued)

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	171.1250000	15.5568182	12.04	<.0001
Error	12	15.5000000	1.2916667		
Corrected Total	23	186.6250000			

R-Square	Coeff Var	Root MSE	y Mean
0.916946	14.43194	1.136515	7.875000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
block	5	108.3750000	21.6750000	16.78	<.0001
A	1	4.0000000	4.0000000	3.10	0.1039
B	1	42.2500000	42.2500000	32.71	<.0001
A*B	1	2.2500000	2.2500000	1.74	0.2115
block*A*B	3	14.2500000	4.7500000	3.68	0.0436

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	5	31.41666667	6.283333333	4.86	0.0116
A	1	4.00000000	4.00000000	3.10	0.1039
B	1	42.25000000	42.25000000	32.71	<.0001
A*B	1	2.25000000	2.25000000	1.74	0.2115
block*A*B	3	14.25000000	4.75000000	3.68	0.0436

Tests of Hypotheses Using the Type III MS for block*A*B as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	4.00000000	4.00000000	0.84	0.4265
B	1	42.25000000	42.25000000	8.89	0.0585
A*B	1	2.25000000	2.25000000	0.47	0.5407

The Mixed Procedure

Model Information

Data Set	WORK.FACTORIAL
Dependent Variable	y
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Table 11.21 (Continued)

Iteration History				
Iteration	Evaluations	-2 Res	Log Like	Criterion
0	1		63.93019485	
1	1		61.40414523	0.00000000

Convergence criteria met.

ANALYSIS OF 2**2 FACTORIAL

The Mixed Procedure

Covariance Parameter
Estimates

Cov Parm	Estimate
block*A*B	1.7292
Residual	1.2917

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
block	5	3	1.32	0.4353
A	1	3	0.84	0.4265
B	1	3	8.89	0.0585
A*B	1	3	0.47	0.5407

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
A	1.0000	1.0897	3	0.92	0.4265
B	3.2500	1.0897	3	2.98	0.0585
A*B	0.7500	1.0897	3	0.69	0.5407

Least Squares Means

Effect	A	B	Estimate	Standard Error	DF	t Value	Pr > t
A	0		7.3750	0.7034	3	10.48	0.0019
A	1		8.3750	0.7034	3	11.91	0.0013
B		0	6.2500	0.7034	3	8.89	0.0030
B		1	9.5000	0.7034	3	13.51	0.0009
A*B	0	0	5.3750	1.0433	3	5.15	0.0142
A*B	0	1	9.3750	1.0433	3	8.99	0.0029
A*B	1	0	7.1250	1.0433	3	6.83	0.0064
A*B	1	1	9.6250	1.0433	3	9.23	0.0027

EXAMPLE 11.11: Using SAS PROC FACTEX we generate the 2^3 factorial design in blocks of size 4 as given in Section 11.6.1 The input statements are given in Table 11.22a. We make the following comments:

- (i) The “factors” statement gives the names of the factors. The default is that these factors have 2 levels.
- (ii) The “blocks” input specifies the size of the blocks.
- (iii) The “model” statement indicates which effects and interactions we want to estimate; in this example we specify the main effects and two-factor interactions as indicated. If we specify too many effects and interactions no suitable system of confounding may exist.

The output is given in Table 11.22b:

- (iv) The design is given in two forms, first in the standard order of the treatment combinations, second (because of the output statement) in the order of the blocks. We note here that the levels of the factors are labeled as -1 and 1 , which corresponds to our notation of 0 and 1 .
- (v) The “Block Pseudo-factor Confounding Rules” gives in general the names of the interactions which were chosen to generate the system of confounding. In our example it is the three-factor interaction ABC . \square

EXAMPLE 11.12: We use SAS PROC FACTEX to generate a system of confounding for the 2^5 factorial in blocks of size 8. The input statements and the output are given in Table 11.23a, b, respectively:

- (i) We denote the factors by f_1, \dots, f_5 .
- (ii) We want to estimate all main effects and two-factor interactions.
- (iii) The interactions to be confounded with blocks are given as $f_2 * f_3 * f_4 * f_5$ and $f_1 * f_4 * f_5$, and consequently their generalized interaction $f_1 * f_2 * f_3$, thus achieving the stated objective. \square

EXAMPLE 11.13: Here we use SAS PROC FACTEX to generate a fractional factorial design. Specifically, we consider the $1/2$ fraction of the 2^3 factorial (see Section 11.7.2). We comment briefly on the input statements and the output given in Table 11.24:

- (i) We have given two equivalent input statements which will generate the same design:

Table 11.22 2³ Factorial Design in Blocks of 4

a) Input statements:

```
proc factex;
factors A B C;
blocks size=4;
model estimate=(A|B|C @2);
examine design confounding;
output out=design blockname=block nvals=(1 2);
title '2**3 FACTORIAL IN BLOCKS OF SIZE 4';
run;
```

```
proc print data=design;
run;
```

b.) Output:

2**3 FACTORIAL IN BLOCKS OF SIZE 4

The FACTEX Procedure

Design Points

Experiment Number	A	B	C	Block
1	-1	-1	-1	1
2	-1	-1	1	2
3	-1	1	-1	2
4	-1	1	1	1
5	1	-1	-1	2
6	1	-1	1	1
7	1	1	-1	1
8	1	1	1	2

Block Pseudo-factor Confounding Rules

[B1] = A*B*C

2**3 FACTORIAL IN BLOCKS OF SIZE 4

Obs	block	A	B	C
1	1	-1	-1	-1
2	1	-1	1	1
3	1	1	-1	1
4	1	1	1	-1
5	2	-1	-1	1
6	2	-1	1	-1
7	2	1	-1	-1
8	2	1	1	1

Table 11.23 2^5 Factorial Design in Blocks of 4

a) Input statements:

```
proc factex;  
factors f1-f5;  
blocks size=8;  
model estimate=(f1|f2|f3|f4|f5 @2);  
examine design confounding;  
title '2**5 FACTORIAL DESIGN IN BLOCKS OF SIZE 8';  
run;
```

b) Output:

2**5 FACTORIAL DESIGN IN BLOCKS OF SIZE 8						
The FACTEX Procedure						
Design Points						
Experiment Number	f1	f2	f3	f4	f5	Block
1	-1	-1	-1	-1	-1	2
2	-1	-1	-1	-1	1	3
3	-1	-1	-1	1	-1	3
4	-1	-1	-1	1	1	2
5	-1	-1	1	-1	-1	1
6	-1	-1	1	-1	1	4
7	-1	-1	1	1	-1	4
8	-1	-1	1	1	1	1
9	-1	1	-1	-1	-1	1
10	-1	1	-1	-1	1	4
11	-1	1	-1	1	-1	4
12	-1	1	-1	1	1	1
13	-1	1	1	-1	-1	2
14	-1	1	1	-1	1	3
15	-1	1	1	1	-1	3
16	-1	1	1	1	1	2
17	1	-1	-1	-1	-1	4
18	1	-1	-1	-1	1	1
19	1	-1	-1	1	-1	1
20	1	-1	-1	1	1	4
21	1	-1	1	-1	-1	3
22	1	-1	1	-1	1	2
23	1	-1	1	1	-1	2
24	1	-1	1	1	1	3
25	1	1	-1	-1	-1	3
26	1	1	-1	-1	1	2
27	1	1	-1	1	-1	2
28	1	1	-1	1	1	3
29	1	1	1	-1	-1	4
30	1	1	1	-1	1	1
31	1	1	1	1	-1	1
32	1	1	1	1	1	4

Block Pseudo-factor Confounding Rules

[B1] = f2*f3*f4*f5
[B2] = f1*f4*f5

Table 11.24 1/2 Fraction of 2³ Factorial

a) Input statements:

```
proc factex;  
factors A B C;  
size design=4;  
model estimate=(A B C);  
examine design aliasing confounding;  
title '1/2 FRACTION OF 2**3 FACTORIAL';  
run;
```

```
proc factex;  
factors A B C;  
size fraction=2;  
model res=3;  
run;
```

b.) Output:

1/2 FRACTION OF 2**3 FACTORIAL

The FACTEX Procedure

Design Points

Experiment Number	A	B	C
1	-1	-1	1
2	-1	1	-1
3	1	-1	-1
4	1	1	1

Factor Confounding Rules

$C = A \times B$

Aliasing Structure

$A = B \times C$
 $B = A \times C$
 $C = A \times B$

The size of the design can be stated either as “size design = # of design points (runs)” or as “size fraction = denominator of fraction”.

The model statement can be given either in terms of effects and interactions to be estimated or in terms of the resolution of the design (see Section 11.7.4).

- (ii) The design points are given in Table 11.24b.
- (iii) The “Factor Confounding Rules” gives an expression equivalent to the defining relationship (see (11.35)). The expression

$$C = A * B$$

can be interpreted as the method of constructing the design: Starting with the full 2^2 factorial with factors A and B the levels of factor C are obtained by multiplying for each run the corresponding levels of A and B .

- (iv) The “Aliasing Structure” is obtained as explained in Section 11.7.3. □

EXAMPLE 11.14: In this example we use SAS PROC FACTEX to combine fractionation and confounding. Specifically, we consider the $1/16$ fraction of the 2^8 factorial in blocks of size 8 in the form of a resolution IV design.

The input statements and the output are given in Table 11.25:

- (i) The 16 design points are given in Table 11.25b, assigning them to two blocks.
- (ii) The factor confounding rules specifies four interactions in the defining relationship, each consisting of four factors, that is,

$$\begin{aligned} I &= f_2 * f_3 * f_4 * f_5 = f_1 * f_3 * f_4 * f_6 \\ &= f_1 * f_2 * f_4 * f_7 = f_1 * f_2 * f_3 * f_8 \end{aligned}$$

to which should be added all their generalized interactions.

- (iii) The block pseudo-factor rule specifies one (“estimable”) four-factor interaction to generate the system of confounding.
- (iv) The alias structure indicates that all main effects are estimable (assuming that three-factor interactions are negligible) and which two-factor interactions are aliased with each other.
- (v) The two-factor interactions indicated by $[B]$ are confounded with blocks since they are aliased with $f_1 * f_2 * f_3 * f_4$ in $[B1]$. □

Table 11.25 Fractional Replication with Confounding

a) Input statements:

```
proc factex;
factors f1-f8;
size design=16;
blocks size=8;
model res=4;
examine design aliasing confounding;
title1 '1/16 FRACTION OF 2**8 FACTORIAL';
title2 'IN BLOCKS OF SIZE 8';
run;
```

b.) Output:

```

1/16 FRACTION OF 2**8 FACTORIAL
IN BLOCKS OF SIZE 8

The FACTEX Procedure

Design Points

Experiment
Number    f1    f2    f3    f4    f5    f6    f7    f8    Block
-----
1         -1    -1    -1    -1    -1    -1    -1    -1     2
2         -1    -1    -1     1     1     1     1    -1     1
3         -1    -1     1    -1     1     1     -1     1     1
4         -1    -1     1     1    -1    -1     1     1     2
5         -1     1    -1    -1     1    -1     1     1     1
6         -1     1    -1     1    -1     1    -1     1     2
7         -1     1     1    -1    -1     1     1    -1     2
8         -1     1     1     1     1    -1    -1    -1     1
9          1    -1    -1    -1    -1     1     1     1     1
10         1    -1    -1     1     1    -1    -1     1     2
11         1    -1     1    -1     1    -1     1    -1     2
12         1    -1     1     1    -1     1    -1    -1     1
13         1     1    -1    -1     1     1    -1    -1     2
14         1     1    -1     1    -1    -1     1    -1     1
15         1     1     1    -1    -1    -1    -1     1     1
16         1     1     1     1     1     1     1     1     2

Factor Confounding Rules

f5 = f2*f3*f4
f6 = f1*f3*f4
f7 = f1*f2*f4
f8 = f1*f2*f3

Block Pseudo-factor Confounding Rules

[B1] = f1*f2*f3*f4
```


Table 11.25 (Continued)

Aliasing Structure

f1	
f2	
f3	
f4	
f5	
f6	
f7	
f8	
f1*f2	= f3*f8 = f4*f7 = f5*f6
f1*f3	= f2*f8 = f4*f6 = f5*f7
f1*f4	= f2*f7 = f3*f6 = f5*f8
[B] = f1*f5	= f2*f6 = f3*f7 = f4*f8
f1*f6	= f2*f5 = f3*f4 = f7*f8
f1*f7	= f2*f4 = f3*f5 = f6*f8
f1*f8	= f2*f3 = f4*f5 = f6*f7

11.12 EXERCISES

11.1 Show that for the 2^4 factorial the main effects and interactions represent a complete set of orthogonal contrasts among the 16 treatment combinations.

11.2 Consider a 2^2 factorial experiment in a randomized complete block design with b blocks.

- (i) Define, in terms of the true treatment effects, the main effects and two-factor interaction.
- (ii) Show that the main effects and the two-factor interaction are orthogonal contrasts among the treatment effects.
- (iii) Suppose each experimental unit has 3 observational units. Give an expression for the variance of the estimators for the main effects and interaction.
- (iv) Outline the ANOVA table for the design with b blocks and 3 observational units per experimental unit giving source of variation, d.f., $E(MS)$, and the F -ratios for testing hypotheses about the main effects and interaction.

11.3 Consider the following block designs with 5 blocks and with treatments having a 2^2 factorial structure (with factors A and B , say):

- (a) randomized complete block design with 2 samples per EU, and 2 measurements per sample;
- (b) generalized randomized block design with 4 replications for each treatment per block;
- (c) generalized randomized block design with 2 replications for each treatment per block, and 2 observations per EU.

For each design:

- (ii) Write out an appropriate linear model.
 - (ii) Outline the ANOVA table, giving sources of variation, d.f. (for the d.f. give numbers not formulas), and $E(\text{MS})$.
 - (iii) Indicate how you would test for main effects and interactions.
 - (iv) Give the variance for \hat{A} and give its estimator, that is, $\widehat{\text{var}}(\hat{A})$.
- 11.4** Consider a 2^2 factorial experiment in a completely randomized design with r replications for each treatment combination. Suppose that for each observation, y , information on a covariate, x , is available.
- (i) Using the supplementary information, give the general expression for the (adjusted) estimator for the main effect A .
 - (ii) Let A_{yy} , B_{yy} , $(AB)_{yy}$ and E_{yy} be the sums of squares for A , B , AB and Error, respectively, in the ANOVA table without the covariate. Using similar notation, give general expressions for the corresponding sums of squares when the covariate is included in the analysis.
- 11.5** Suppose you are consulted to help design an experiment involving two factors at two levels each. A sufficient number of blocks of size two are available for the experiment. The investigator wishes to obtain equal information on the main effects and the two-factor interaction.
- (i) Give the name of the method used for constructing a suitable experimental design.
 - (ii) Write out explicitly the design for this study and explain how you obtained it.
 - (iii) Outline the ANOVA table for the design given in (ii), including source of variation, d.f., and sums of squares.
 - (iv) Suppose only 6 blocks of size 2 are available for the experiment. What method could one use to construct the design. Explain and give the design. What kind of design is this?
- 11.6** A horticultural experiment conducted in a green house was laid out as a Latin square design, where the blocking factors represent temperature and light intensity, respectively. The treatments have a 2^2 factorial structure, that is, 2 factors A and B each at 2 levels. The layout of the design and the results from the experiment (in parentheses) are given below:

Temperature	Light Intensity			
	1	2	3	4
1	a_0b_0 (5)	a_1b_1 (10)	a_0b_1 (8)	a_1b_0 (7)
2	a_1b_1 (12)	a_1b_0 (8)	a_0b_0 (6)	a_0b_1 (10)
3	a_1b_0 (10)	a_0b_1 (8)	a_1b_1 (15)	a_0b_0 (7)
4	a_0b_1 (9)	a_0b_0 (9)	a_1b_0 (11)	a_1b_1 (16)

- (i) Give a linear model for analyzing the data from this experiment and sketch the ANOVA giving sources of variation and d.f.
- (ii) Obtain the ANOVA table.
- (iii) Give a numerical expression for the estimate of the interaction $A \times B$.
- (iv) The experiment is to be repeated at different times, so that in the end data from 3 different times, T_1, T_2, T_3 say, will be available (the experiment above represents T_1). Even though the temperature and light intensity trends remain, they may be assumed to differ from one time period to the next. It is expected that there is interaction between the time factor and the treatment factors.

Give a linear model for data from this experiment and sketch the ANOVA table, giving sources of variation and d.f.

- (v) For the experiment described in (iv) what is the variance of the estimated main effects, \hat{A} , \hat{B} , and interaction, \widehat{AB} ?

11.7 Suppose a dermatologist wants to study the effectiveness of two (2) different preparations of a skin lotion using two (2) different forms of application (for example, one vs. two applications per day). He has available 12 patients with a certain skin disease and he can apply one form of medication (that is, combination of preparation and frequency of application) to each arm of each patient. Even though the patients have the same disease, there exists considerable variation among them, but the two arms of a patient are quite homogeneous.

- (i) What type of experimental design would be appropriate for this study?
- (ii) What are the experimental units?
- (iii) Give a suitable experimental plan for this study and describe how you obtained this plan.
- (iv) For the design given in (iii), outline the ANOVA table, giving sources of variation, d.f., and sums of squares.

- (v) For the plan given in (iii), what is the variance of the estimates of the main effects and the interaction?

11.8 Consider a 2^6 factorial experiment and suppose that the experimenter has only enough resources to handle just a fraction of all possible treatment combinations. Let this fraction be chosen by the identify relationship

$$I = +ABCD = +CDEF = +ABEF$$

- (i) Give the treatment combinations that make up this fraction.
 - (ii) Assuming that all interactions involving 3 or more factors and all 2-factor interactions *not* involving factor B can be estimated from this fractional factorial.
 - (iii) Suppose we have 2 replications (that is, 2 EUs) for each treatment combination in a CRD; outline the ANOVA table (giving sources of variation, d.f. and $E(MS)$) based on the assumptions given in (ii).
 - (iv) Suppose we need to use blocks of size 8 and we have 4 blocks available; under the assumptions in (ii) give a suitable arrangement without sacrificing information about the main effects and 2-factor interactions involving factor B.
 - (v) For the design in (iv) outline the ANOVA table, giving sources of variation and d.f.
 - (vi) Describe how you would obtain the ANOVA table in (v) with SAS.
- 11.9** Use SAS PROC FACTEX to construct designs equivalent to those given in (i) and (iv) in Exercise 11.8.
- 11.10** Show that for the 3^3 factorial contrasts belonging to A and ABC^2 are orthogonal to each other.
- 11.11** Construct a system of partial confounding for the 3^2 factorial in blocks of size 3 with 6 blocks so that at least partial information can be obtained about the 2-factor interaction components and full information about the main effects.
- 11.12** Obtain the treatment combinations and the alias structure for a $1/3$ fraction of the 3^4 factorial (assuming that 3- and 4-factor interactions are negligible).
- 11.13** For the fraction obtained in Exercise 11.12, obtain a system of confounding using blocks of size 9. State what assumptions need to be made for this design to be useful.
- 11.14** Consider the $2^3 \times 3^2$ factorial. Obtain a suitable system of confounding for blocks of size 6.

This Page Intentionally Left Blank

CHAPTER 12

Response Surface Designs

12.1 INTRODUCTION

We have mentioned earlier that much of the topic on experimental design, and certainly most of this book, is concerned with what we call comparative experiments. The emphasis and, in fact, the whole purpose of the experiment here is the comparison of treatments. We have explored this topic in detail in Chapter 7 for the CRD, with obvious extensions to other error-control designs. In our discussion we have distinguished between qualitative and quantitative treatments, but in both cases the aim has been the same: to detect structure of some form among the treatment effects. In the case of quantitative treatments this can be done by using methods of regression analysis. If, for example, a straight line (with a nonzero slope) can be fitted to characterize the dependence of the estimated treatment effects on the treatments then this tells us not only that the treatment effects are different from each other, but also that there exists a simple relationship among them.

More generally, the dependence of treatment effects on treatments can be represented as a response curve (if the treatments are represented by the levels of one treatment factor, for example, amount of fertilizer) or a response surface (if the treatments are level combinations of two or more treatment factors, for example, amount of fertilizer and rate of application). And such curves or surfaces can be used to make judgments not only about treatment structure but also about the relationship between treatments and responses, or between input variables and output variables. Knowledge of this relationship is important if one wants, for example, to find the treatment combination which gives the optimal (highest or lowest) response. We shall never know the exact relationship but we can try to approximate it. This is done, often sequentially, by using methods of experimental design and regression analysis. Methods that are directed towards this kind of investigation, using tools from experimental design and regression analysis, are usually referred to as *response surface methodology* (RSM).

RSM was developed mainly with a view towards industrial experimentation and production (see Box and Wilson, 1951) but it has found application also in agriculture (see Mead and Pike, 1975), in medical settings (see Carter, Wampler, and Stablein,

1983), and more recently in connection with off-line quality control (see Vining and Myers, 1990). And even though RSM has proven to be useful in practice, it suffers from a serious defect, namely that the form of the response surface depends on the choice of units for the input variables. To illustrate this point we consider a simple example. The relationship $y = x_1^2 + x_2^2$ can be pictured as a two-dimensional surface in a three-dimensional space giving the dependence of y on x_1 and x_2 . If we change the units for the input variables to $x_1^* = 2x_1$ and $x_2^* = 3x_2$, then the relationship becomes

$$y = \frac{1}{4}x_1^{*2} + \frac{1}{9}x_2^{*2}.$$

While y is constant on the curves $x_1^2 + x_2^2 = \text{constant}$, that is, on circles in the (x_1, x_2) -plane, it is constant on the curves $\frac{1}{4}x_1^{*2} + \frac{1}{9}x_2^{*2} = \text{constant}$, that is, on ellipses in the (x_1^*, x_2^*) -plane. Obviously, these surfaces are quite different from each other illustrating the point that there is a surface only with a choice of units of plotting, a point that must be kept in mind in the following discussion.

12.2 FORMULATION OF THE PROBLEM

Suppose we have k quantitative factors F_1, F_2, \dots, F_k which are known or suspected to have an effect on a particular response. Each factor has continuous levels within a certain interval; for example, F_i has levels X_i with $X_{iL} \leq X_i \leq X_{iU}$ ($i = 1, 2, \dots, k$). The hypercube $\{X_{iL} \leq X_i \leq X_{iU}; i = 1, 2, \dots, k\}$ contains the so-called *operational region* (OR) in which every level combination (X_1, X_2, \dots, X_k) is a feasible operating condition. We assume that each such setting can be controlled (essentially without error) by the experimenter. To each setting (X_1, X_2, \dots, X_k) belongs a response, η , which is some function of the levels, that is,

$$\eta = \phi(X_1, X_2, \dots, X_k; \theta_1, \theta_2, \dots, \theta_q), \quad (12.1)$$

where $\theta_1, \theta_2, \dots, \theta_q$ are parameters. We write (12.1) for short as

$$\eta = \phi(\mathbf{X}; \boldsymbol{\theta}) \quad (12.2)$$

with $\mathbf{X} = (X_1, X_2, \dots, X_k)'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)'$. Now both the true yield, $\eta = \eta(X_1, X_2, \dots, X_k)$, at any given point in OR and the form of the functional relationship ϕ are unknown. Instead, we will have available only observed responses $y = y(\mathbf{X})$ and we shall attempt to approximate $\phi(\mathbf{X}, \boldsymbol{\theta})$ by a polynomial function $f(\mathbf{X}, \boldsymbol{\beta})$ in \mathbf{X} . We then consider in place of (12.2) a model of the form

$$y(\mathbf{X}) = f(\mathbf{X}, \boldsymbol{\beta}) + e(\mathbf{X}), \quad (12.3)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)'$ are unknown parameters and $e(\mathbf{X})$ represents error.

Ideally we would like to have $y(\mathbf{X})$ available for a sufficiently fine grid in OR in order to approximate ϕ , or rather a realization of ϕ , sufficiently well. From a practical point of view this is clearly impossible. Instead we will be restricted to a relatively small number of points (these are sometimes referred to as runs or experiments) which will typically be confined to a region which is called the *experimental region* (ER) or

region of interest. Obviously, ER is contained in OR. The basic idea, due to Box and Wilson (1951), here is then the following: Based on our limited knowledge about the process under study we choose an ER. We assume that the response surface for ER is sufficiently smooth and hence can be approximated by a low-order polynomial, for example, of first or second degree. We then choose an appropriate treatment and error-control design to estimate the coefficients of the polynomial. From this we predict the response for any point in ER. If one of these points attains the optimal response then presumably our goal is achieved (we may, however, have reached a local optimum rather than a global optimum); if the fitted response surface indicates that the optimum may be outside ER then we would have to choose a new ER and repeat the whole process until the (predicted) optimum can be located.

As Box and Wilson (1951) point out, the procedure described above leads to two sources of error: (i) experimental and sampling error in estimating the function $f(\mathbf{X}; \beta)$ of (12.3) and (ii) bias due to the inadequacy of $f(\mathbf{X}; \beta)$ approximating $\phi(\mathbf{X}; \theta)$ of (12.2). To minimize these errors, singly or jointly, is essentially the focus of response surface designs. To this end, Box and Hunter (1957), suggested the following basic requirements for such designs:

- (i) Assuming that a polynomial $f(\mathbf{X}; \beta)$ of degree d approximates $\phi(\mathbf{X}; \theta)$ sufficiently well, the design should allow $f(\mathbf{X}; \beta)$ to be estimated with satisfactory precision in ER.
- (ii) The design should allow to check whether the chosen $f(\mathbf{X}; \beta)$ provides a satisfactory fit to the response surface or whether a different polynomial may be more appropriate.
- (iii) The design should not contain an excessively large number of experimental points.
- (iv) The design should lend itself to adequate blocking of the experimental points.
- (v) One should be able to amend the design in case the polynomial of degree d proves to be inadequate and a polynomial of degree $d + 1$ needs to be fitted.

These requirements were refined and expanded by Box (1968) and Box and Draper (1975) (see also Box and Draper, 1987). We shall not go into the details here but rather concentrate in the following on the five fundamental points above.

In the following we shall describe the basic tools and designs of RSM and point out connections to treatment and error-control designs discussed elsewhere in this book. For details and further developments of RSM we refer the reader to specialized texts on this subject, for example, Box and Draper (1987), Khuri and Cornell (1996), Myers and Montgomery (2002).

12.3 FIRST-ORDER MODELS AND DESIGNS

12.3.1 First-Order Regression Model

Within a small region it is often not unreasonable to approximate the response surface, that is, the function ϕ , by a first-order polynomial in the k input variables X_1, X_2, \dots, X_k :

$$y = \beta_0 + \sum_{i=1}^k \beta_i X_i + e. \quad (12.4)$$

In this model the regression coefficient β_i is a measure of the change in the response y due to a change in the input variable X_i . This is, of course, the kind of information provided by the main effects from a factorial experiment where each factor has two levels (see Section 11.3). A natural choice of a response surface design for this situation is therefore a 2^k factorial, or a fraction of it.

Suppose then we have 2^k experimental points $(X_1, X_2, \dots, X_k)_j$ say, with $j = 1, 2, \dots, 2^k$. With each level combination being replicated r times in a CRD we have $N = r2^k$ experimental runs. Denote the low and high level of the i th factor by X_{i0} and X_{i1} , respectively. If we use instead of X_i the coded levels

$$x_i = \frac{X_i - \bar{X}}{\frac{1}{2}(X_{i1} - X_{i0})} \quad (12.5)$$

then the low and high levels become $x_{i0} = -1$ and $x_{i1} = 1$, respectively. We rewrite (12.4) now as

$$y(x_1, x_2, \dots, x_k)_l = \beta_0^* + \sum_{i=1}^k \beta_i^* x_i + e(x_1, x_2, \dots, x_k)_l, \quad (12.6)$$

where $x_i = \pm 1$, or in matrix notation as

$$\mathbf{y} = (\mathbf{J}, \mathbf{D})\boldsymbol{\beta}^* + \mathbf{e},$$

where \mathbf{y} is the $N \times 1$ vector of observations, \mathbf{J} is an $N \times 1$ vector of unity elements, \mathbf{D} is the $N \times k$ design-model matrix of -1 's and 1 's, $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, \beta_2^*, \dots, \beta_k^*)'$, and \mathbf{e} is the $N \times 1$ vector of errors. More specifically, if we write

$$\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_k)$$

as k $N \times 1$ column vectors, we know that each \mathbf{d}_i has $r2^{k-1}$ elements equal to -1 and $r2^{k-1}$ elements equal to 1 . This implies that $\mathbf{J}'\mathbf{d}_i = 0$ for every i . Moreover, we have $\mathbf{d}_i'\mathbf{d}_{i'} = 0$ for every i, i' with $i \neq i'$, that is, the \mathbf{d}_i 's are orthogonal to each other.

12.3.2 Least Squares Analysis

Using the properties given above, the normal equations for the β_i^* of (12.6),

$$(\mathbf{J}, \mathbf{D})'(\mathbf{J}, \mathbf{D})\hat{\boldsymbol{\beta}}^* = (\mathbf{J}, \mathbf{D})'\mathbf{y}$$

simplify to

$$N\hat{\beta}^* = \begin{pmatrix} \mathbf{J}'\mathbf{y} \\ \mathbf{D}'\mathbf{y} \end{pmatrix}. \quad (12.7)$$

It then follows that

$$\hat{\beta}_0^* = \frac{1}{N} \sum_{x_1, x_2, \dots, x_k} \sum_l y(x_1, x_2, \dots, x_k)_l = \bar{y}$$

and

$$\begin{aligned} \hat{\beta}_i^* &= \frac{1}{N} \mathbf{d}_i' \mathbf{y} \\ &= \frac{1}{N} [(\text{sum of all observations with } x_i = 1) \\ &\quad - (\text{sum of all observations with } x_i = -1)]. \end{aligned} \quad (12.8)$$

We note that $\hat{\beta}_i^*$ is half the corresponding main effect given in Section 11.3. It follows further from (12.7) that

$$\text{var}(\hat{\beta}_i^*) = \frac{1}{N} \sigma_e^2 \quad (12.9)$$

and

$$\text{cov}(\hat{\beta}_i^*, \hat{\beta}_l^*) = 0.$$

Hence, for any given point $\mathbf{z} = (z_1, z_2, \dots, z_k)'$ in the ER given by $\{-1 \leq z_i \leq 1; i = 1, 2, \dots, k\}$ we obtain the predicted response

$$\hat{y}(\mathbf{z}) = \hat{\beta}_0^* + \sum_{i=1}^k \hat{\beta}_i^* z_i \quad (12.10)$$

with

$$\text{var}[\hat{y}(\mathbf{z})] = \frac{1}{N} \left(1 + \sum_{i=1}^k z_i^2 \right) \sigma_e^2. \quad (12.11)$$

In order to evaluate which factors are influential and to investigate the response surface [as given by the $\hat{y}(\mathbf{z})$] in more detail we need to obtain an estimate of σ_e^2 . This is achieved, as usual, through the ANOVA as given in Table 12.1, where

$$y(\mathbf{x})_l = y(x_1, x_2, \dots, x_k)_l$$

$$\sum_x = \sum_{x_1, x_2, \dots, x_k}$$

$$D_1 = \text{SS}(\text{Total}) - N \sum_{i=1}^k (\hat{\beta}_i^*)^2$$

$$D_2 = D_1 - \text{SS}(PE)$$

Table 12.1 ANOVA for First-Order Response Surface Design

Source	d.f.	SS
Regression	k	
β_1^*	1	$N(\hat{\beta}_1^*)^2$
β_2^*	1	$N(\hat{\beta}_2^*)^2$
\vdots	\vdots	\vdots
β_k^*	1	$N(\hat{\beta}_k^*)^2$
Error	$r2^k - k - 1$	$D_1 = \text{SS}(E)$
Lack-of-fit error	$2^k - k - 1$	$D_2 = \text{SS}(LOF)$
Pure error	$2^k(r - 1)$	$\sum_x \sum_l (y(\mathbf{x})_l - \bar{y}(\mathbf{x}))^2 = \text{SS}(PE)$
Total	$N - 1$	$\sum_{x,l} (y(\mathbf{x})_l - \bar{y}(\cdot))^2$

We mention here that $\text{SS}(E)$ consists of two parts, the usual error sum of squares for a CRD, denoted here by $\text{SS}(PE)$, that is, sum of squares for pure error, and the sum of the sums of squares for all interactions for the 2^k factorial denoted here by $\text{SS}(LOF)$. As in regression analysis this sum of squares can be used to test whether the postulated model (12.6) provides a sufficiently good enough fit to the data, a point to which we shall return later. To test whether the i th factor contributes to explaining the response we use the F -test

$$F_i = \frac{\text{SS}(\beta_i^*)}{\text{MS}(E)} \quad (i = 1, 2, \dots, k)$$

with 1 and $\nu = N - k - 1$ d.f. Suppose we consider, without loss of generality, only the first k_1 factors to be important. We may then reconsider model (12.6) and use instead

$$y(x_1, x_2, \dots, x_k)_l = \beta_0^* + \sum_{i=1}^{k_1} \beta_i^* x_i + e(x_1, x_2, \dots, x_k)_l \quad (12.12)$$

and

$$\hat{y}(\mathbf{z}) = \hat{\beta}_0^* + \sum_{i=1}^{k_1} \hat{\beta}_i^* z_i \quad (12.13)$$

with

$$\widehat{\text{var}}[\hat{y}(\mathbf{z})] = \frac{1}{N} \left(1 + \sum_{i=1}^{k_1} z_i^2 \right) \text{MS}(E). \quad (12.14)$$

We may then compare the responses for two different sets of input variables, say $\mathbf{z} = (z_1, z_2, \dots, z_{k_1})'$ and $\mathbf{w} = (w_1, w_2, \dots, w_{k_1})'$, by considering

$$\hat{y}(\mathbf{z}) - \hat{y}(\mathbf{w}) = \sum_{i=1}^{k_1} \hat{\beta}_i^* (z_i - w_i) \quad (12.15)$$

and

$$\widehat{\text{var}}[\hat{y}(\mathbf{z}) - \hat{y}(\mathbf{w})] = \frac{1}{N} \sum_{i=1}^{k_1} (z_i - w_i)^2 \text{MS}(E). \quad (12.16)$$

Similarly, we may consider differences in responses if some of the input variables are kept constant at a desired level and the remaining input variables are varied to achieve optimum response if indeed it can be achieved in ER. Due to the fact that we are approximating the true response surface and due to experimental error there may, of course, not exist a single level combination which achieves the optimum response but rather the estimated responses in the neighborhood of an optimum may not be significantly different from each other.

12.3.3 Alternative Designs

The use of a full 2^k factorial to estimate the parameters of a first-order response surface will usually be wasteful, especially if it is used in a CRD with r replications for each level combination. There are basically two ways to reduce the number of experimental points. One way is to replicate each design point (x_1, x_2, \dots, x_k) only once, that is, $r = 1$. In that case we have $\text{SS}(PE) = 0$ and $\text{SS}(E) = \text{SS}(LOF)$ (see Table 12.1). Another way is to use only a fraction of a 2^k factorial (see Section 11.7) either as a single replicate or as a CRD with $r > 1$ replications. In either case we need to choose a fraction such that all k main effects are estimable and that sufficient d.f. for error will be available so that comparisons of the type (12.15) can be made with satisfactory statistical power as measured by the variance (12.16). This means that if we were to choose a very small fraction, such as a resolution III fractional factorial, we need several replications for each design point. Even if we were to choose a fractional factorial of resolution IV or V we may need some replication. Methods for constructing fractional factorials are discussed in Chapters II.13 and 14.

An important property of a 2^k factorial is that blocking can be accommodated easily without sacrificing estimation of the main effects, that is, the β_i^* . Such blocking may become necessary for a number of reasons, mainly determined by practical and experimental considerations. For example, it may not be possible to complete all experimental runs with one batch of raw material and one suspects systematic batch-to-batch variation. For a full factorial appropriate blocks can be obtained (by using the methods indicated in Section 11.6 and more fully developed in Chapter II.8) as long as the block size, 2^l , is larger than the number of factors, k , for example, a 2^3 in blocks of size 4, a 2^4 in blocks of size 8, a 2^5 in blocks of size 8 or 16, and so on. Similar blocking arrangements can also be constructed for fractions of a 2^k factorial, for example, a $1/2$ fraction of a 2^5 in blocks of size 8, a $1/4$ fraction of a 2^6 in blocks of size 8, and so on.

An alternative to the factorial designs described above is a class of designs referred to as *simplex designs* (Box, 1952). These are orthogonal designs, that is, the columns of the design-model matrix \mathbf{D} satisfy $\mathbf{d}_i' \mathbf{d}_{i'} = 0$ for $i \neq i'$ with $k + 1$ design points, which are then replicated r times in a CRD or a RCBD. The design points are located at the vertices of a regular k -dimensional simplex, which for $k = 2$ is an equilateral triangle, for $k = 3$ is a tetrahedron, and so on. For $r = 1$ the matrix \mathbf{D} can in general be written as (see Khuri and Cornell, 1996)

$$\mathbf{D} = \begin{pmatrix} -a_1 & -a_2 & -a_3 & \dots & -a_k \\ a_1 & -a_2 & -a_3 & \dots & -a_k \\ 0 & 2a_2 & -a_3 & & \vdots \\ 0 & 0 & 3a_3 & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & -a_k \\ 0 & 0 & 0 & \dots & ka_k \end{pmatrix},$$

where $a_i = c_i[(k + 1)/i(i + 1)]^{1/2}$ and c_i are scaling factors (it is common practice to choose $c_i = c$ for every i). Since the simplex design contains only $k + 1$ experimental points to estimate $k + 1$ regression coefficients it has zero d.f. for $SS(LOF)$ (see Table 12.1).

12.4 SECOND-ORDER MODELS AND DESIGNS

12.4.1 Second-Order Linear Regression

One advantage of using a 2^k factorial with r replications over a simplex design with r replications is that the factorial design provides an opportunity to check the adequacy of the model (12.4) through the F -test

$$F_{LOF} = \frac{MS(LOF)}{MS(PE)}$$

(see Table 12.1). This test allows us to check whether interactions among the factors are present and if so the design enables us to obtain the various sums of squares for two-factor interactions, three-factor interactions, and so on (see Section 11.3). This, however, may provide only part of the answer, why (12.4) is not a good approximation to the true response surface $\phi(X_1, X_2, \dots, X_k)$. Another reason for an inadequate fit may be that curvature due to various factors is present. This, however, cannot be detected with a 2^k experiment. We shall now consider an extension of model (12.4) incorporating some form of curvature and interaction and suitable designs to estimate the parameters of such models.

Table 12.2 ANOVA for Second-Order Response Surface Design

Source	d.f.
Regression	$2k + \frac{1}{2}k(k-1)$
Linear effects	k
Quadratic effects	k
Linear \times linear effects	$\frac{1}{2}k(k-1)$
Error	$r3^k - 2k - \frac{1}{2}k(k-1) - 1$
Lack of fit	$3^k - 2k - \frac{1}{2}k(k-1) - 1$
Pure	$3^k(r-1)$

A second-order model for k input variables is defined as

$$y(X_1, X_2, \dots, X_k) = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \beta_{ii} X_i^2 + \sum_{i < j} \beta_{ij} X_i X_j + e(X_1, X_1, \dots, X_k) \quad (12.17)$$

12.4.2 Possible Designs

An obvious but usually not the best design for estimating the parameters of this model is a 3^k factorial. If we choose the levels for each factor to be equidistant we can reparameterize (12.17) in terms of orthogonal polynomials (see Section 11.9.1) and obtain estimates of the linear, quadratic and linear \times linear effects for all factors and two-factor combinations, that is, of α_{pq} with $p, q = 0, 1, 2$ and $p + q \leq 2$ (see model (11.49). For r replications of each design point in a CRD a sketch of the ANOVA is given in Table 12.2 (for other details see Table 11.16). Here again we can partition the d.f. for error into d.f. for lack of fit (accounting for interactions other than linear \times linear) and d.f. for pure error (arising from replications). Even for small k the number of d.f. for lack of fit is quite substantial, stemming from a large number of experimental points and the (assumed) absence of most interaction effects. There are several ways in which we can reduce the excessive number of experimental points:

- (i) We can eliminate replication of experimental points, that is, choose $r = 1$.
- (ii) We can use fractional factorials with or without replication, but we have to restrict ourselves now to resolution V designs so that main effects (linear and quadratic) and two-factor interactions can be estimated, for example a $1/3$ fraction of a 3^5 or 3^6 , or a $1/9$ fraction of a 3^7 (see Sections 11.9.5, 11.9.6, and Chapter II.13). Even in these cases the number of experimental points is generally excessive for the number of parameters (regression coefficients) to be estimated.

- (iii) We may try to construct designs which are more suited for the specific situation with a limited number of design points. Several such classes of designs have been proposed. We shall consider briefly two—*central composite designs* and *Box-Behnken designs*.

12.4.3 Central Composite Designs

The *central composite design* (CCD) was introduced by Box and Wilson (1951). Each factor is used at five different levels, but not all level combinations occur. Rather, the CCD is composed of three parts:

- (a) a factorial or “cube” part consisting of 2^{k-p} points from a full 2^k factorial ($p = 0$) or a $1/2^p$ fraction of the 2^k factorial of at least resolution V (see above), each point being replicated r_f times; the levels of each factor are coded as -1 and $+1$; the number of experimental runs is $n_f = 2^{k-p}r_f$;
- (b) an axial or “star” part consisting of $2k$ points on the axis of each factor at a distance α from the center of the design, each point being replicated r_a times; this gives rise to $n_a = 2kr_a$ experimental runs;
- (c) n_0 replications of the center point $(0, 0, \dots, 0)$.

The total number of experimental runs then is $N = n_f + n_a + n_0$.

EXAMPLE 12.1: For $k = 2$, the basic CCD is as given in Figure 12.1. The design matrix for this design, \mathbf{D}^* say, with $r_f = 1$, $r_a = 1$, $n_0 = 1$ can be written as

$$\mathbf{D}^* = \begin{pmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ \alpha & 0 \\ -\alpha & 0 \\ 0 & \alpha \\ 0 & -\alpha \\ 0 & 0 \end{pmatrix}. \quad (12.18)$$

□

The values for α , r_f , r_a , and n_0 can be chosen to obtain certain properties of the design and to satisfy economic requirements. One such property, that of rotatability, was introduced by Box and Hunter (1957). A design is said to be *rotatable* if the prediction variance, for a level combination $\mathbf{z} = (z_1, z_2, \dots, z_k)'$ in ER, that is, $\text{var}[\hat{y}(\mathbf{z})]$, is the same for all points that are equidistant from the design center. This property is satisfied for the first-order designs discussed in Section 12.3 (see (12.11) which depends only on $\sum z_i^2$) and it is satisfied simply because the columns of \mathbf{D} are orthogonal (and hence the design is orthogonal) and because of the scaling used. For second-order designs the conditions are more complex in general having to do with the so-called design mo-

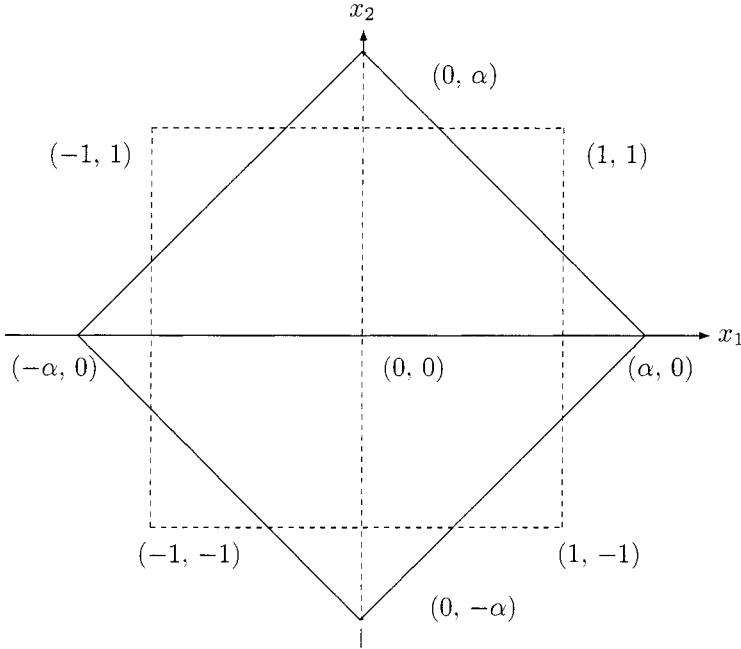


Figure 12.1 CCD for $k = 2$.

ments [see Box and Hunter (1957), Box and Draper (1996), Khuri and Cornell (1996), Myers and Montgomery (2002)]. For the CCD the conditions are satisfied by choosing

$$\alpha = \left(\frac{n_f}{r_a} \right)^{1/4}. \quad (12.19)$$

For example, for $k = 2, r_f = r_a$, we have $\alpha = \sqrt{2}$.

12.4.4 Blocking in Central Composite Designs

An important property for CCDs is that of orthogonal blocking as discussed by Box and Hunter (1957). This idea is similar to that of systems of confounding in factorial experiments (see Section 11.6) which also leads to orthogonal blocking, except that we have to deal here with the different components of the CCD. If we denote the levels of the i th factor for the N experimental runs by x_{il} ($l = 1, 2, \dots, N$), each x_{il} being one of $-\alpha, -1, 0, 1, \alpha$, then Box and Hunter (1957) give the following two conditions for orthogonal blocking: (i) Each block must itself be a first-order orthogonal design, and (ii) the fraction of the total sum of squares of each input variable contributed by every block must equal the fraction of the total observations in the block. Suppose we have b blocks and the size of the u th block is n_u , that is, $\sum_{a=1}^b n_u = N$. Then, according to

(i) we must have

$$\sum_{l(u)} x_{il} x_{jl} = 0 \quad (i, j = 0, 1, \dots, k; i \neq j) \quad (12.20)$$

for every $u = 1, 2, \dots, b$, where $\sum_{l(u)}$ means summation over all l in block u . Condition (ii) can be written as

$$\frac{\sum_{l(u)} x_{il}^2}{N} = \frac{n_u}{N} \quad (i = 1, 2, \dots, k) \quad (12.21)$$

for every $u = 1, 2, \dots, b$.

Different blocking schemes satisfying (12.20) and (12.21) can be derived from considering first the case of $b = 2$ blocks. One block consists of all n_f runs from the factorial part plus n_{0f} center runs; the other block consists of all n_a runs from the axial part plus n_{0a} center runs, where $n_{0f} + n_{0a} = n_0$. It is obvious by just looking at \mathbf{D}^* that condition (12.20) is satisfied for all pairs (i, j) and both blocks. Condition (12.21) for the first block is

$$\frac{n_f}{n_f + 2r_a\alpha^2} = \frac{n_f + n_{0f}}{N} \quad (12.22)$$

and for the second block

$$\frac{2r_a\alpha^2}{n_f + 2r_a\alpha^2} = \frac{n_a + n_{0a}}{N}, \quad (12.23)$$

which is the same for every $i = 1, 2, \dots, k$. Combining (12.22) and (12.23) yields

$$\frac{n_f}{2r_a\alpha^2} = \frac{n_f + n_{0f}}{n_a + n_{0a}}. \quad (12.24)$$

Thus (12.24) gives us the value of α such that (12.21) holds, that is,

$$\begin{aligned} \alpha^2 &= \frac{kn_f}{n_a} \cdot \frac{n_a + n_{0a}}{n_f + n_{0f}} \\ &= k \frac{1 + n_{0a}/n_a}{1 + n_{0f}/n_f}. \end{aligned} \quad (12.25)$$

Typically n_{0a}/n_a and n_{0f}/n_f are quite small, so that $\alpha \cong \sqrt{k}$ which means that the axial points have about the same distance from the center as the factorial points.

Orthogonal blocking with smaller blocks than those discussed above can be obtained, using similar arguments, in a number of ways. We mention just a few:

- (i) If $r_f > 1$, each block may consist of one or more replicates of all 2^k points plus some center runs.
- (ii) Systems of confounding as discussed in Section 11.6 and in Chapters II.8 and 9 may be used together with some center runs.

- (iii) If $r_a > 1$, each set of axial points plus some center runs may form a block.
- (iv) One block may consist of half of the axial points, say all points with $+\alpha$, plus some center runs, and the other blocks have all axial points with $-\alpha$ plus some center runs.

Methods (i) and (ii) may be combined with methods (iii) and (iv). Similar methods can be used if a fractional factorial is used for the factorial part of the CCD, remembering that those fractions have to be at least of resolution V .

Box and Draper (1987) give some recommendations for the choice of r_f and r_a , and Draper (1982) discusses criteria for deciding on $n_0 = n_{0f} + n_{0a}$, the number of center runs. If more experimental runs are needed, for example to increase d.f. for pure error, it is often convenient to increase n_0 without sacrificing other properties of the CCD.

12.4.5 Box-Behnken Designs

We have mentioned earlier that using a 3^k factorial for a second-order response surface usually results in too many experimental points. The CCDs discussed above correct this situation, but they use five levels for each factor. An interesting class of designs using only three levels of each factor and at the same time resulting in a “reasonable” number of experimental points was proposed by Box and Behnken (1960). These designs can be constructed by combining ideas from incomplete block designs (BIBD or PBIBD; see Section 9.8 and Chapters II.1–5) and factorial experiments, specifically 2^k factorials. The method can be described as follows.

Suppose we have t input variables x_1, x_2, \dots, x_t and an incomplete block design with t treatments and b blocks of size k . This design is characterized by its incidence matrix $\mathbf{N} = (n_{li})$ with $n_{li} = 1$ if treatment l occurs in block i and $n_{li} = 0$, otherwise. We now identify the t treatments with the t input variables and consider \mathbf{N}' . Each row of \mathbf{N}' contains k unity elements. Suppose in the first row they occur in columns l_1, l_2, \dots, l_k . We then replace these k unity elements successively by the level combinations of the 2^k factorial where the k factors are the input variables l_1, l_2, \dots, l_k . The $t - k$ zeros in the first row are replaced by $2^k \times 1$ vectors of zeros. This procedure is repeated for each row of \mathbf{N}' resulting in $b2^k$ experimental points to which we add n_0 center runs (see Jo and Hinkelmann, 1993).

EXAMPLE 12.2: Consider the case $t = 6$. The matrix \mathbf{N}' of a PBIBD (design R42 in Clatworthy, 1973) with blocks of size $k = 3$ is given by

$$\mathbf{N}' = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Table 12.3 Box-Behnken Design For $t = 6$

x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
-1	-1	0	-1	0	0	-1	0	0	-1	-1	0
1	-1	0	-1	0	0	1	0	0	-1	-1	0
-1	1	0	-1	0	0	-1	0	0	1	-1	0
1	1	0	-1	0	0	1	0	0	1	-1	0
-1	-1	0	1	0	0	-1	0	0	-1	1	0
1	-1	0	1	0	0	1	0	0	-1	1	0
-1	1	0	1	0	0	-1	0	0	1	1	0
1	1	0	1	0	0	1	0	0	1	1	0
0	-1	-1	0	-1	0	0	-1	0	0	-1	-1
0	1	-1	0	-1	0	0	1	0	0	-1	-1
0	-1	1	0	-1	0	0	-1	0	0	1	-1
0	1	1	0	-1	0	0	1	0	0	1	-1
0	-1	-1	0	1	0	0	-1	0	0	-1	1
0	1	-1	0	1	0	0	1	0	0	-1	1
0	-1	1	0	1	0	0	-1	0	0	1	1
0	1	1	0	1	0	0	1	0	0	1	1
0	0	-1	-1	0	-1	-1	0	-1	0	0	-1
0	0	1	-1	0	-1	1	0	-1	0	0	-1
0	0	-1	1	0	-1	-1	0	1	0	0	-1
0	0	1	1	0	-1	1	0	1	0	0	-1
0	0	-1	-1	0	1	-1	0	-1	0	0	1
0	0	1	-1	0	1	1	0	-1	0	0	1
0	0	-1	1	0	1	-1	0	1	0	0	1
0	0	1	1	0	1	1	0	1	0	0	1
						0	0	0	0	0	0

The level combinations for the 2^3 factorial are

x_{l_1}	x_{l_2}	x_{l_3}
-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
1	-1	1
-1	1	1
1	1	1

Substituting x_{l_1} , x_{l_2} , and x_{l_3} for the 1's in each row and adding one center run we obtain the Box-Behnken design of Table 12.3. □

To obtain economical designs we need to choose incomplete block designs with k and b small so that $N = b2^k + n_0$ does not become too large. For larger k we may use a fractional factorial of resolution V instead of the complete factorial. For particular choices of incomplete block designs (that is, resolvable designs, see Chapter II.2) orthogonal blocking is possible. This may also be achieved by using a system of confounding such that no main effects and no two-factor interactions are confounded with blocks. Box and Behnken (1960) give a list of designs for some values of t together with possibilities for orthogonal blocking.

12.4.6 Hard-to-Change versus Easy-to-Change Factors

We have mentioned earlier that the response surface designs, for instance, the CCD of (12.18), are embedded in some form of error control design, in most cases a CRD or RCBD. This means, of course, that the treatment combinations are randomly assigned to the experimental units. In industrial experimentation it is often the case that the runs are performed sequentially. Random assignment then means random order of application. This implies that the factors have to be reset for each run, whether there is a level change or not. In practice the resetting is often not done if a factor remains the same between two or more runs, either because of convenience or because the factor in question is hard to change. This has led to the notion of *hard-to-change factors* (HTC) and *easy-to-change factors* (ETC).

Webb, Lucas and Borkowski (2004) report on an example with one HTC factor and two ETC factors:

EXAMPLE 12.3: An experiment was performed to investigate three factors in the operation of a wrapper machine: spacing of the seal crimper, speed of the machine, temperature of the seal crimper. Spacing was recognized as a HTC factor, whereas speed and temperature were thought to be ETC factors. The experiment, using a Box-Behnken design (see Section 12.4.5), was set up in “blocks” of levels of the HTC factor, that is, in each “block” the HTC factor, was at the same level and no resetting took place. Within the “blocks” the ETC factor levels were randomized according to the chosen design, except when the experiment was actually performed it was found that speed also turned out to be a HTC factor. As a consequence, this factor was not reset when the same level occurred in consecutive runs. As a result, the experiment was conducted as illustrated in Table 12.4, where the lines indicate the “blocks” of not reset levels for spacing and speed. \square

We shall not pursue this example here further, referring the reader to Webb, Lucas and Borkowski (2004), except to say that this is an example of a split-split-plot type experiment (see Section 13.6) which is highly unbalanced and generally undesirable. Because of the split-split-plotting two additional errors will be induced and, as a consequence, an analysis using generalized least squares (GLS) (see Section 4.16.2) needs to be performed instead of ordinary least squares (OLS) with model (12.17).

To avoid some of the complications associated with such an unbalanced design it is advisable to construct designs that show a certain amount of balance. In addition it

Table 12.4 Wrapper Machine Example

Spacing	Speed	Temp
0	1	-1
0	1	1
0	0	0
1	0	1
1	0	-1
1	-1	0
1	1	0
-1	1	0
-1	-1	0
-1	0	-1
-1	0	1
0	0	0
0	-1	-1
0	-1	1
0	1	1

would be desirable to be able to estimate the parameters in the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{e}_A + \mathbf{e}_B, \quad (12.26)$$

such that the estimates are identical under GLS and OLS. In (12.26) \mathbf{X} and \mathbf{Z} are known matrices, where \mathbf{Z} is determined by “blocking” of the HTC factors. In the terminology of split-plot designs (see Chapter 13) the HTC factors are called whole-plot factors and the ETC factors are the split-plot factors. Thus, \mathbf{e}_A and \mathbf{e}_B are referred to as whole-plot and split-plot errors, respectively.

Parker, Kowalski and Vining (2007) refer to designs that satisfy the properties mentioned above as *equivalent estimation designs*. They provide techniques for constructing such designs, of which the following is an example.

EXAMPLE 12.4: Suppose we want to use a CCD for two factors A and B say, as given in (12.18). We identify A as a HTC factor and B as a ETC factor. An equivalent estimation CCD consists then of $r(\geq 2)$ replicates of the design given in Table 12.5.

For this design we have for \mathbf{Z} in (12.26)

$$\mathbf{Z} = \begin{pmatrix} \mathbf{J}_2 & & & & \\ & \mathbf{J}_2 & & & 0 \\ & & \mathbf{J}_2 & & \\ & & & \mathbf{J}_2 & \\ 0 & & & & \mathbf{J}_2 \\ & & & & & \mathbf{J}_2 \end{pmatrix}$$

□

If there are more than one HTC factors, then each of these factors individually is kept at the same level for each whole plot. For example, if we have two HTC factors, A_1 and A_2 , two whole plots of the design may look as follows:

Whole plot	A_1	A_2
1	1	-1
	1	-1
2	-1	-1
	-1	-1

It is important to remember that A_2 has to be reset in whole plot 2 even though the levels are the same in whole plots 1 and 2.

12.5 INTEGRATED MEAN SQUARED ERROR DESIGNS

In our discussion of first-order and second-order designs we have made the assumption that the first-order and second-order models, respectively, are satisfactory approximations to the true response surfaces. If this is true then the designs discussed are appropriate. Often, however, there is the fear that the assumption may not be right. We refer to such a situation as *model misspecification*. We may suspect, for example, that instead of a first-order model a second-order model may provide a better approximation to the true situation, but we are not sure. An obvious reaction would be to use a second-order design so that we can estimate all second-order effects. The drawback of this approach is that if our suspicion is not true then we have wasted valuable resources by using too many experimental points. One must, therefore, find some compromise for the choice of an appropriate design; firstly, it must enable estimation of the parameters of the specified model sufficiently well; secondly, it must provide some protection

Table 12.5 Basic Replicate of Equivalent Estimation CCD

A	B
-1	-1
-1	1
1	-1
1	1
α	0
α	0
$-\alpha$	0
$-\alpha$	0
0	α
0	$-\alpha$
0	0
0	0

against model misspecification; and thirdly, it must be economical. We shall explain this for a very simple situation and then make some more general comments.

12.5.1 Variance and Bias for the One-Factor Case

Consider the case of one factor X and suppose we approximate $\phi(X, \theta)$ by

$$f(X, \beta) = \beta_0 + \beta_1 X$$

Suppose further that ER is defined by $X_L \leq X \leq X_U$, or in the coded variable x by $-1 \leq x \leq 1$. For a given set of N x -values, x_1, x_2, \dots, x_N with $\bar{x} = 0$, we then fit the model

$$y = \beta_0^* + \beta_1^* x + e \quad (12.27)$$

and obtain the predicted (estimated) response curve

$$\hat{y}(z) = \hat{\beta}_0^* + \hat{\beta}_1^* z \quad (12.28)$$

for any z in $[-1, 1]$. Denote the true value of the response curve at z by $\phi(z)$. Then the mean squared error associated with estimating $\phi(z)$ by $\hat{y}(z)$ is given by

$$\begin{aligned} E[\hat{y}(z) - \phi(z)]^2 &= E\{\hat{y}(z) - E[\hat{y}(z)] + E[\hat{y}(z)] - \phi(z)\}^2 \\ &= \text{var}[\hat{y}(z)] + \{E[\hat{y}(z)] - \phi(z)\}^2. \end{aligned} \quad (12.29)$$

For purposes of comparison it is useful to normalize (12.29) for the number of experimental points, N , and the error variance, σ_e^2 , as

$$\frac{NE[\hat{y}(z) - \phi(z)]^2}{\sigma_e^2} = \frac{N \text{var}[\hat{y}(z)]}{\sigma_e^2} + \frac{N\{E[\hat{y}(z)] - \phi(z)\}^2}{\sigma_e^2}, \quad (12.30)$$

which we write for short as

$$M(z) = V(z) + B(z), \quad (12.31)$$

where $V(z)$ is the variance at z and $B(z)$ is the squared bias at z . For the special case here we find [see (12.11)]

$$V(z) = 1 + \frac{z^2}{\sigma_x^2}, \quad (12.32)$$

where $\sigma_x^2 = (1/N)\sum_l x_l^2$. σ_x^2 is also referred to as the second design moment and denoted by $[1 \quad 1]$ (analogously, $(1/N)\sum x_l$ is referred to as the first design moment, denoted by $[1]$, which equals zero in our case, and $[1 \quad 1 \quad 1] = (1/N)\sum x_l^3$ is the third design moment). The bias portion of $M(z)$ depends, of course, on $\phi(z)$. Suppose that

$$E[y(x)] = \phi(x) = \beta_0^* + \beta_1^*x + \beta_2^*x^2. \quad (12.33)$$

In order to evaluate $E[\hat{y}(z)]$ we shall write (12.33) in matrix notation as

$$E(\mathbf{y}) = (\mathbf{X}_1 : \mathbf{X}_2) \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix},$$

where

$$\mathbf{X}_1 = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ x_N^2 \end{bmatrix}$$

$$\gamma_1 = \begin{pmatrix} \beta_0^* \\ \beta_1^* \end{pmatrix}, \quad \gamma_2 = \beta_2^*.$$

Now

$$\begin{aligned} \hat{\mathbf{y}}(z) &= (1, z)\hat{\gamma}_1 \\ &= (1, z)(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{y} \end{aligned}$$

and

$$\begin{aligned}
 E[\hat{y}(z)] &= (1, z)(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 E(\mathbf{y}) \\
 &= (1, z)(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 (\mathbf{X}_1; \mathbf{X}_2) \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \\
 &= (1, z)(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 (\mathbf{X}_1 \gamma_1 + \mathbf{X}_2 \gamma_2) \\
 &= \beta_0^* + \beta_1^* z + (1, z)(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \gamma_2
 \end{aligned} \tag{12.34}$$

$$= \beta_0^* + \beta_1^* z + \left\{ \frac{1}{N} \sum_{l=1}^N x_l^2 + \frac{z \sum x_l^3}{\sum x_l^2} \right\} \beta_2^*. \tag{12.35}$$

In (12.34), the matrix $(\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2$ is called the *alias matrix*. Using the moment notation, $B(z)$ can now be written as

$$B(z) = \frac{N\beta_2^{*2}}{\sigma_e^2} \left\{ \frac{1}{[1 \quad 1]} ([1 \quad 1]^2 + [1 \quad 1 \quad 1]z) - z^2 \right\}^2. \tag{12.36}$$

Rather than consider the mean squared error just for an arbitrary point, z , it is more informative to consider some sort of average mean squared error, referred to as *integrated mean squared error* (IMSE) and defined as

$$M = \frac{\int_{\text{ER}} M(z) dz}{\int_{\text{ER}} dz}. \tag{12.37}$$

Performing this operation for $V(z)$ and $B(z)$, we obtain for the IMSE

$$\begin{aligned}
 M &= \frac{\int_{\text{ER}} [V(z) + B(z)] dz}{\int_{\text{ER}} dz} \\
 &= V + B, \text{ say.}
 \end{aligned} \tag{12.38}$$

For our example we have $\text{ER} = [-1, 1]$ and hence $\int_{\text{ER}} dz = \int_{-1}^1 dz = z$. Substituting (12.32) and (12.36) in (12.38), we obtain

$$M = \left\{ 1 + \frac{1}{3[1 \quad 1]} \right\} + \frac{N\beta_2^{*2}}{\sigma_e^2} \left\{ [1 \quad 1]^2 + \frac{[1 \quad 1 \quad 1]^2}{3[1 \quad 1]^2} - \frac{2[1 \quad 1]}{3} + \frac{1}{5} \right\}. \tag{12.39}$$

One would like to choose a design that minimizes M and from (12.39) it can be seen that this has to be done by affecting the design moments $[1 \quad 1]$ and $[1 \quad 1 \quad 1]$. Unfortunately, such a choice will also be influenced by the unknown parameter β_2^*/σ_e , the standardized measure of the true curvature. Furthermore, the choice of the design will depend on the relative magnitude of V and B and how important they appear, relative to each other, to the investigator.

12.5.2 Choice of Design

There are two extreme cases: (i) V is much larger than B and (ii) B is much larger than V . For all practical purposes we can think of (i) as $V > 0, B = 0$. To minimize M then means to minimize V and that, in turn, means to maximize $[1 \ 1]$. For N even, this is achieved by choosing $N/2$ experimental runs at $x = -1$ and $x = +1$, respectively. For N odd, we choose $(N - 1)/2$ runs at $x = -1$ and $x = +1$ each and one run at $x = 0$, with $[1 \ 1] = (N - 1)/N$. Such designs (which assume $B = 0$) are referred to as *all-variance designs*.

Case (ii) above can be characterized essentially by $V = 0, B > 0$. To minimize M then is to minimize B . The first step might be to choose a design with $[1 \ 1 \ 1] = 0$. It follows then from (12.39) that now

$$B = \frac{N\beta_2^{*2}}{\sigma_e^2} \left\{ \left([1 \ 1] - \frac{1}{3} \right)^2 + \frac{4}{45} \right\}$$

and hence minimizing B requires a design with $[1 \ 1] = \frac{1}{3}$. Such a design (which assumes $V = 0$) is referred to as an *all-bias design*. Comparing the values for $[1 \ 1]$ for the all-variance and all-bias designs shows that the spread of the experimental points for the all-bias design is much smaller than that for the all-variance designs. In fact, the conditions for $[1 \ 1]$ for these two types of design are in conflict with each other and hence minimizing M cannot be achieved by minimizing V and B separately.

For the general problem of minimizing M one may start again by choosing $[1 \ 1 \ 1] = 0$ and then minimize the resulting expression for M with respect to $[1 \ 1]$ for different values of $N\beta_2^{*2}/\sigma_e^2$. Alternatively, one can minimize M with respect to $[1 \ 1]$ as a function of $N\beta_2^{*2}/\sigma_e^2$ and then choose the value of $[1 \ 1]$ which forces V/B to be a certain value α which expresses the experimenter's opinion about the relative values of V and B . For example, $\alpha = 1$ implies that V and B are equally important. In that case the optimal value of $[1 \ 1]$ is .388 (for $\sqrt{N}\beta_2^{*2}/\sigma_e = 4.49$), comparable to that for the all-bias design (Box and Draper, 1959). Computations for other cases show that when curvature is suspected the optimal design is closer to the all-bias design than to the all-variance design (Box and Draper, 1959; Khuri and Cornell, 1996).

It is apparent from this simple example that choosing a design which minimizes IMSE is rather complex and becomes even more so if we consider a first-order model or a second-order model and want to protect against second-order effects or third-order effects, respectively. General considerations for a specific class of designs indicate, however, that optimal designs tend to be close to all-bias designs. More specific results are provided by Box and Draper (1963) (see also Box and Draper, 1987, and Khuri and Cornell, 1996). It must be emphasized, however, that many of these results are somewhat subjective in that they are not always invariant to the scaling of the input variables. Hence those results may be taken as general guidelines only when choosing an appropriate design.

12.6 SEARCHING FOR AN OPTIMUM

As we have pointed out earlier, RSM is a sequential process based on subject matter and statistical input. Experiments are performed using the investigator's best knowledge about the process under study and the statistician's recommendations how to best perform the experiment (see also Chapter 2). After having decided which factors (input variables) should be studied and in which range of levels, usually a first-order design is used to approximate the response surface in the chosen ER. In the search for an optimum response and the levels of the factors at that optimum response one can imagine many scenarios leading to a sequence of experiments and statistical decisions. It is, of course, impossible to describe every situation that might possibly arise, instead we shall mention briefly some of the steps in the sequence of events.

Following each experiment it is important to study the estimated response surface in some detail. This requires that the underlying design has been chosen with those goals in mind. We may, for example, want to

- (i) investigate which factors are important,
- (ii) examine whether the chosen polynomial provides an adequate approximation to the response surface,
- (iii) plot the contours of the response surface,
- (iv) decide on a new ER,
- (v) locate the optimum response as quickly as possible.

Some of these goals are relatively easy to obtain by using designs of the type we have discussed in Sections 12.3 and 12.4. We can test hypotheses about the regression coefficients in the model to check (i). Assuming that the design chosen allows the estimation of pure error (constituting experimental and observational error) or if such information is available from other sources, we can examine (ii) through a lack-of-fit test as exemplified in the ANOVAs given in Tables 12.1 and 12.2. But even drawing a contour map, that is, a map of equal responses, $\hat{y}(\mathbf{x})$, for different input variables \mathbf{x} is not always easy. Even if we could draw in a k -dimensional space the shape of the contours depend crucially on the scaling used for the input variables.

The reason why we mention this is the fact that, loosely speaking, the contour map is used to locate new ERs in the pursuit of locating the optimum response. Different mathematical techniques have been proposed and are used to find the most direct path to the optimum. They all depend on the contour map as established from the results of the initial experiment and updated by subsequent experiments as determined by the optimization procedure used.

The procedure most often discussed is the *method of steepest ascent* which was introduced in RSM by Box and Wilson (1951). (For a detailed discussion see also Box and Draper, 1987, and Khuri and Cornell, 1996). A direction perpendicular to the contour planes as established by a first-order model or contour surfaces for a second-order model is the direction of steepest ascent, pointing towards higher responses. Along this path further experiments are performed until a best value or apparent maximum

is reached. Such a point may serve as the center point for a new ER in which then a more comprehensive experiment will be performed, continuing this cycle as long as necessary, changing most likely from first-order designs to second-order designs as the situation warrants. The virtues and value of the method of steepest ascent have been put into question when Johnson in the discussion of the Box and Wilson (1951) paper pointed out that the method suffers from dependence on the scale (that is, choice of units) of the input variables. As a consequence, a certain amount of care must be used when applying it. Subjective scaling will lead to subjective directions of experimentation and only through checks can a potentially misleading direction be avoided. A perhaps more useful method would be to scale the input variables such that the change of one unit for one variable is as important as the change of one unit in another variable. But even that is not entirely objective and may depend on the location in the OR.

To avoid the problem of scale-dependence, other optimization procedures have been proposed. The *method of parallel tangents* (PARTAN) was introduced by Shah, Buehler, and Kempthorne (1964) and further discussed by Buehler, Shah, and Kempthorne (1964). Another approach, using simplex designs, was proposed by Spendley, Hext, and Hinsworth (1962). Any discussion of these methods is beyond the scope of this chapter and the reader is referred to the pertinent literature.

12.7 EXPERIMENTS WITH MIXTURES

12.7.1 Defining the Problem

A special and yet quite distinct application of response surface methodology occurs in experiments with mixtures. The special feature of these experiments is that the response (12.1) does not depend on the actual values (amounts) of the input (represented by the input variables X_1, X_2, \dots, X_k) but rather on the proportions relative to each other, that is, for a mixture of three ingredients we might have $X_1 = 50\%$, $X_2 = 25\%$, $X_3 = 25\%$ with, of course, $X_1 + X_2 + X_3 = 100\%$. An example of such a mixture experiment may be the blending of three gasoline stocks to determine the blend which will give the best mileage.

The pioneering work in this area was done by Scheffé (1958) who introduced simplex-lattice designs and appropriate polynomial models to investigate the type of question mentioned above. An excellent account of current methodology and thinking is given by Cornell, (2002). We shall give here only a very brief discussion of some of the design aspects in this area, to what extent they are different from designs for comparative experiments and to what extent they make use of the designs we have discussed in other chapters. For details the reader should refer to Cornell, (2002) and the references therein.

Let X_1, X_2, \dots, X_k be the input variables which are constrained by the condition

$$\sum X_i = 1 \quad (12.40)$$

This condition introduces a dependence among the X_i 's which means that they cannot take on all values in R_k^+ but only in the k -dimensional simplex. The coordinate system used for these values is called the *simplex coordinate system*. For $k = 3$, for example,

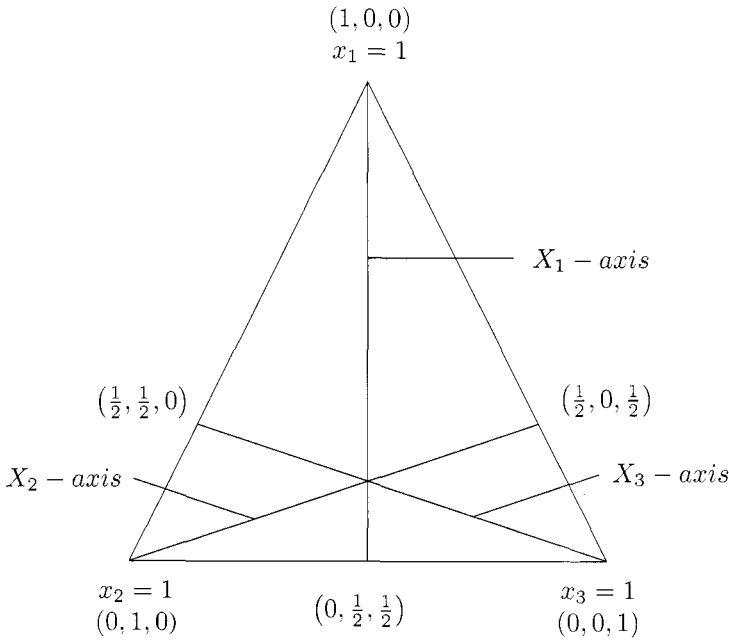


Figure 12.2 Triangular Coordinate Paper.

this coordinate system can be plotted on triangular graph paper with lines parallel to the sides of the equilateral triangle as given in Figure 12.2

In order to fit an approximate model of the form (12.3), usually a linear or quadratic model, we need to conduct an experiment in order to obtain appropriate observations, $y(\mathbf{X})$. Three types of designs are used most often: *simplex-lattice designs*, *simplex-centroid designs*, and *axial designs*.

12.7.2 Simplex-Lattice Designs

The name *simplex-lattice design* refers to a collection of uniformly spaced points on a simplex. For a k -dimensional simplex there exist different simplex-lattice designs depending on the spacing of the levels, that is, the proportions for each component X_i ($i = 1, 2, \dots, k$) may take the $m + 1$ equally spaced values

$$X_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1$$

subject to (12.40). Such a lattice is referred to as a (k, m) lattice. For example, the $(3, 2)$ lattice consists of the points

$$(X_1, X_2, X_3) = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (\frac{1}{2}, \frac{1}{2}, 0), (\frac{1}{2}, 0, \frac{1}{2}), (0, \frac{1}{2}, \frac{1}{2})\}$$

and these points lie all on the vertices and sides of the simplex, that is, triangle (see Figure 12.2). For the (3, 3) lattice the points are

$$(X_1, X_2, X_3) = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (\frac{2}{3}, \frac{1}{3}, 0), (\frac{2}{3}, 0, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}, 0), (\frac{1}{3}, 0, \frac{2}{3}), (0, \frac{2}{3}, \frac{1}{3}), (0, \frac{1}{3}, \frac{2}{3}), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\}.$$

Here we have in addition to points on the boundary of the simplex also a design point at the centroid of the simplex.

12.7.3 Simplex-Centroid Designs

For a k -component *simplex-centroid design*, the design points are such that either one, or two, or three, ..., or k components are included in the mixture and if l ($1 \leq l \leq k$) components are included in the mixture they are included in equal proportions, that is, $1/l$. Thus the simplex-centroid design consists of $2^k - 1$ points: k permutations of $(1, 0, \dots, 0)$; $\binom{k}{2}$ permutations of $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$, ..., $\binom{k}{l}$ permutations of $(\frac{1}{l}, \frac{1}{l}, \dots, \frac{1}{l}, 0, \dots, 0)$, ..., and the centroid $(\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k})$. The points are located at the centroid of the $(k - 1)$ -dimensional lattice and at the centroids of all lower-dimensional simplexes contained in the $(k - 1)$ -dimensional simplex.

12.7.4 Axial Designs

Whereas for the simplex-lattice design and the simplex-centroid design, the design points (with the exception of the overall centroid) are located on the boundaries of the simplex, the design points for the *axial design* are located on the component axes (see Figure 12.2 for $k = 3$). This implies that for every such point all k components are included in the mixture. A simple form of such a design was suggested by Cornell (1975). In it the points are located at equal distances, say $\Delta_1, \Delta_2, \dots$, from the centroid $(\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k})$ toward each of the vertices.

12.7.5 Canonical Polynomials

For all three types of designs, and combinations of them, the number of points (runs) depends to some extent on the degree of the polynomial to be fitted to the data. Typically, the polynomials are of the first or second degree, that is,

$$y(\mathbf{X}) = \beta_0 + \sum_{i=1}^k \beta_i X_i + e \quad (12.41)$$

or

$$y(\mathbf{X}) = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \beta_{ii} X_i^2 + \sum_{i < j}^k \beta_{ij} X_i X_j + e. \quad (12.42)$$

The number of points, obviously, has to be at least as large as the number of parameters to be estimated in (12.41) or (12.42) or any other model that might be appropriate. We

could proceed then as usual, except that for the situation described in this section we have to take condition (12.40) into account and because of it the parameters associated with the various X -terms are not unique. To remove the dependence among the X -values we could write, for example,

$$X_k = 1 - \sum_{i=1}^{k-1} X_i \quad (12.43)$$

and substitute (12.43) in (12.41) and (12.42). In (12.41) this would lead to a model with parameters $\beta_1 - \beta_k$ ($i = 1, 2, \dots, k-1$) obscuring the separate effects of the individual components. The effects on model (12.42) are even more complex.

The preferred way of removing the effect of the dependence among the X_i is to obtain the so-called *canonical polynomial*. For the linear polynomial this form is obtained by multiplying β_0 in (12.41) by $\sum X_i (= 1)$ and then simplifying, that is,

$$y(\mathbf{X}) = \beta_0 \left(\sum_{i=1}^k X_i \right) + \sum_{i=1}^k \beta_i X_i + e = \sum_{i=1}^k \beta_i^* X_i + e \quad (12.44)$$

with $\beta_i^* = \beta_0 + \beta_i$ ($i = 1, 2, \dots, k$). Model (12.44) retains the symmetry in the k components and the β_i^* have a clear meaning.

To achieve the canonical form for model (12.42) we proceed in the same way as above and use in addition

$$X_i^2 = X_i \left(1 - \sum_{\substack{j=1 \\ j \neq i}}^k X_j \right).$$

Collecting terms leads to the canonical model

$$y(\mathbf{X}) = \sum_{i=1}^k \beta_i^* X_i + \sum_{i < j}^k \sum_{i < j} \beta_{ij}^* X_i X_j + e \quad (12.45)$$

with $\beta_i^* = \beta_0 + \beta_i + \beta_{ii}$ and $\beta_{ij}^* = \beta_{ij} - \beta_{ii} - \beta_{jj}$ ($i, j = 1, 2, \dots, k; i < j$). Model (12.45) can be simplified still further by multiplying $\sum \beta_i^* X_i$ by $\sum X_i$ which yields

$$y(\mathbf{X}) = \sum_{i \leq j}^k \delta_{ij} X_i X_j + e \quad (12.46)$$

with $\delta_{ii} = \beta_i^*$ and $\delta_{ij} = \beta_{ij}^* + \beta_i^* + \beta_j^*$ ($i, j = 1, 2, \dots, k; i < j$). Models (12.45) and (12.46) are, of course, equivalent and contain the same number of parameters.

With data obtained from an appropriate design, models (12.44) or (12.45) or (12.46) can be fitted using the method of least squares. Appropriate tests of hypotheses including lack of fit can then be performed in the usual fashion. From the estimated regression coefficients the response surface can be predicted and standard errors can be obtained using familiar procedures (for details see Cornell, 2002).

12.7.6 Including Process Variables

So far we have discussed the situation where the response η of (12.1) depends only on the mixture variables X_i . In many practical situations there may, however, be other variables, not connected with the blending process itself, which influence η . Such variables are referred to as *process variables*. In the example of blending gasoline stocks such process variables may, for example, be type of car (light, heavy) and speed of driving (slow, fast). If we denote the process input variables by Z_1, Z_2, \dots, Z_p then model (12.1) may be generalized to

$$\eta(\mathbf{X}, \mathbf{Z}) = \phi(X_1, X_2, \dots, X_k; Z_1, Z_2, \dots, Z_p; \theta_1, \theta_2, \dots, \theta_s) \quad (12.47)$$

and (12.47) will be approximated by a (low degree) polynomial in X , Z , and XZ . This will be used to assess not only the effects of the blending variables (X), but also the additive effects of the various “levels” of the process variables (Z) and, quite importantly, the possible interactions (XZ) between the blending and process variables.

The added problem then is to augment the design for the blending variables (as discussed above) with a design for the process variables. The latter will typically be a factorial design. In its simplest form this may be a 2^p factorial or, in order to keep the number of runs at a reasonable level, it may be a fractional factorial (see Sections 11.6 and 11.7). The total number of runs is then determined by the number of design points for the mixture experiment and the number of treatment combinations used for the process variables in that each mixture experiment is performed at each process variable combination included in the factorial or fractional factorial design. To make the entire experiment more manageable the device of blocking may have to be used. Also, it may be possible to reduce the number of runs by combining the design points for the two component designs (that is, blending and process) in a way different from that described above. This provides an example how mixture designs and designs for comparative experiments can be combined in a useful way, and how elements from response surface methodology and design of comparative experiments can be brought to bear on problems that arise in several types of industries, such as chemical, food, textile industries and others.

12.8 EXAMPLES USING SAS®

Since analyzing data from response surface experiments involves regression models the most appropriate SAS procedure to use for the analysis are PROC REG or PROC RSREG. However, for certain purposes and in certain situations, also PROC GLM and PROC MIXED prove to be useful. We shall illustrate the use of these procedures in the following examples.

EXAMPLE 12.5: Consider a first-order design for $k = 3$ variables with the treatment combinations for the 2^3 factorial as the design points, each replicated twice in a CRD. The design and the observations are given in Table 12.6a.

For the analysis we use PROC REG to estimate and test the regression coefficients. The results are given in Table 12.6b, with all three regression coefficients significant at $P < .0001$.

In order to obtain explicitly $SS(LOF)$ we use PROC GLM with x_1, x_2, x_3 as classification variables. In addition to specifying in the model statement x_1, x_2, x_3 we also include the 3-factor interaction term $x_1 * x_2 * x_3$. This is a device for collecting all interaction terms, three 2-factor and one 3-factor interaction, into one sum of squares, which constitutes $SS(LOF)$ with 4 d.f. of Table 12.1. The P -value 0.4233 indicates that there is no lack of fit.

Finally, we point out that the test statistics for testing significance of the regression coefficients are not identical since in the first analysis $[SS(LOF) + SS(PE)]/12 = 0.1725$ is used as the error term, whereas in the second analysis $SS(PE)/8 = 0.1675$ is used.

EXAMPLE 12.6: Consider the CCD in two variables as given in Table 12.7a. To analyze the data we use PROC RSREG. The output is given in Table 12.7b. We make the following comments:

- (i) Simply inputting the two variables, A and B, in the model statement leads to a second-order model and analysis.
- (ii) Including the option “lackfit” in the model statement leads to a partitioning

$$SS(E) = SS(LOF) + SS(PE)$$

The results indicate that there is no lack of fit ($P = .42$).

- (iii) The first-order regression coefficients are significant ($P = .05$, and $.04$, respectively), whereas the second-order coefficients are not significant with $P = .15$, $.14$, and $.11$, respectively.

EXAMPLE 12.7: Consider the CCD as given in Table 12.8a and two situations under which this experiment could have been performed: (a) as a CRD in which case the block classification is ignored, or (b) as a split-plot type design, where factor A is the hard-to-change factor (see Section 12.4.6) and we have “blocks” of size 2.

We comment on both analyses as given in Table 12.8b:

CRD:

- (i) The number of d.f. for error equals 6 with 3 d.f. due to LOF and 3 d.f. due to pure error, and $MS(E) = 1.05$.
- (ii) The linear regression coefficients are significant with $P = .016$ and $.013$, respectively, whereas the quadratic and mixed regression coefficients are marginally significant with $P = .061, .064, .087$, respectively.

Table 12.6 First-order Design and Analysis

a) Input statements:

```
data first;
input x1 x2 x3 y @@;
datalines;

-1 -1 -1 10.1 -1 -1 -1 11.3
 1 -1 -1 12.0  1 -1 -1 11.7
-1  1 -1 13.2 -1  1 -1 12.9
 1  1 -1 14.5  1  1 -1 14.7
-1 -1  1 13.4 -1 -1  1 13.9
 1 -1  1 15.3  1 -1  1 14.9
-1  1  1 16.6 -1  1  1 16.0
 1  1  1 18.2  1  1  1 18.7
;
run;

proc reg data=first;
model y = x1 x2 x3;
title1 'FIRST-ORDER DESIGN';
title2 'REGRESSION ANALYSIS';
run;

proc glm data=first;
class x1 x2 x3;
model y = x1 x2 x3 x1*x2*x3/ss3;
title2 'TESTING FOR LACK OF FIT';
run;
```

b) Output:

FIRST-ORDER DESIGN					
REGRESSION ANALYSIS					
The REG Procedure					
Model: MODEL1					
Dependent Variable: y					
Number of Observations Read				16	
Number of Observations Used				16	
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	84.94750	28.31583	164.15	<.0001
Error	12	2.07000	0.17250		
Corrected Total	15	87.01750			
Root MSE		0.41533	R-Square	0.9762	
Dependent Mean		14.21250	Adj R-Sq	0.9703	
Coeff Var		2.92230			

Table 12.6 (Continued)

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	14.21250	0.10383	136.88	<.0001
x1	1	0.78750	0.10383	7.58	<.0001
x2	1	1.38750	0.10383	13.36	<.0001
x3	1	1.66250	0.10383	16.01	<.0001
FIRST-ORDER DESIGN TESTING FOR LACK OF FIT					
The GLM Procedure					
Class Level Information					
Class	Levels	Values			
x1	2	-1 1			
x2	2	-1 1			
x3	2	-1 1			
Number of Observations Read				16	
Number of Observations Used				16	
Dependent Variable: y					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	85.67750000	12.23964286	73.07	<.0001
Error	8	1.34000000	0.16750000		
Corrected Total	15	87.01750000			
R-Square	Coeff Var	Root MSE	y Mean		
0.984601	2.879632	0.409268	14.21250		
Source	DF	Type III SS	Mean Square	F Value	Pr > F
x1	1	9.92250000	9.92250000	59.24	<.0001
x2	1	30.80250000	30.80250000	183.90	<.0001
x3	1	44.22250000	44.22250000	264.01	<.0001
x1*x2*x3	4	0.73000000	0.18250000	1.09	0.4233

Table 12.7 Regression Analysis for CCD

a) Input statements:

```
data second;
input A B y;
datalines;

-1 -1 3.0
-1 1 4.5
1 -1 4.1
1 1 9.8
1.4 0 6.7
-1.4 0 4.8
0 1.4 7.0
0 -1.4 5.0
0 0 3.0
0 0 4.1
;
run;

proc rsreg data=second;
model y=A B/lackfit predict;
title1 'CENTRAL COMPOSITE DESIGN';
title2 'REGRESSION ANALYSIS';
run;
```

b) Output:

CENTRAL COMPOSITE DESIGN
REGRESSION ANALYSIS

The RSREG Procedure

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
A	0	1.400000
B	0	1.400000

Response Surface for Variable y

Response Mean	5.200000
Root MSE	1.146902
R-Square	0.8666
Coefficient of Variation	22.0558

Regression	DF	Type I Sum of Squares	R-Square	F Value	Pr > F
Linear	2	22.990354	0.5829	8.74	0.0347
Quadratic	2	6.778105	0.1719	2.58	0.1910
Crossproduct	1	4.410000	0.1118	3.35	0.1411
Total Model	5	34.178459	0.8666	5.20	0.0677

Residual	DF	Sum of Squares	Mean Square	F Value	Pr > F
Lack of Fit	3	4.656541	1.552180	2.57	0.4233
Pure Error	1	0.605000	0.605000		
Total Error	4	5.261541	1.315385		

Table 12.7 (Continued)

Parameter	DF	Estimate	Standard Error	t Value	Pr > t	Parameter Estimate from Coded Data
Intercept	1	3.561442	0.810817	4.39	0.0118	3.561442
A	1	1.143939	0.407534	2.81	0.0485	1.601515
B	1	1.262626	0.407534	3.10	0.0363	1.767677
A*A	1	0.970668	0.543320	1.79	0.1485	1.902509
B*A	1	1.050000	0.573451	1.83	0.1411	2.058000
B*B	1	1.098219	0.543320	2.02	0.1133	2.152509

Factor	DF	Sum of Squares	Mean Square	F Value	Pr > F
A	3	18.972482	6.324161	4.81	0.0817
B	3	22.410532	7.470177	5.68	0.0633

Predicted value at stationary point: 3.097108

Stationary point is a minimum.

Split-plot:

- (iii) The estimates of the regression parameters are the same as for the CRD, but the standard errors are different: they are larger for A , $A * A$, and $B * B$ and smaller for B and $A * B$ (see (iv) below).
- (iv) Since factor A is the whole-plot factor, the regression coefficients A and $A * A$ are evaluated against the whole-spot error MS, which is equal to $.8472 + 2 \times .3065$ as obtained from the covariance parameter estimates. Factor B is the split-plot factor and hence the regression coefficient B is evaluated against the split-plot error, which is $.8472$. However, the regression coefficient $B * B$ is confounded with the whole-plot (because the squared values of the levels for factor B do not change within a whole-plot), hence the larger standard error. The regression coefficient $A * B$ is associated with the split-plot and hence has a smaller standard error.
- (v) The comments about the various standard errors in (iv) are also reflected in the different d.f. associated with the tests about the regression coefficients. We have 2 d.f. for the whole-plot error and 4 d.f. for the split-plot error.
- (vi) Only the regression coefficient B is clearly significant ($P = .02$), whereas A and $A * B$ are marginally significant ($P = .11$) and $P = .08$, respectively). \square

Table 12.8 Central Composite Design

a) Input statements:

```
data ccd;
input A B block y;
datalines;
-1 -1 1 13.0
-1 1 1 14.5
1 -1 1 24.1
1 1 1 29.8
1.4 0 3 6.2
1.4 0 3 7.1
-1.4 0 4 4.9
-1.4 0 4 4.6
0 1.4 5 7.0
0 -1.4 5 5.0
0 0 6 3.0
0 0 6 4.1
;
run;

proc glm data=ccd;
model y= A B A*A B*B A*B/solution;
title1 'CENTRAL COMPOSITE DESIGN';
title2 'AS COMPLETELY RANDOMIZED DESIGN';
run;
```

```
proc mixed data=ccd;
class block;
model y=A B A*A B*B A*B/solution ddfm=Satterth;
title2 'AS SPLIT-PLOT DESIGN';
random block;
run;
```

b) Output:

CENTRAL COMPOSITE DESIGN
AS COMPLETELY RANDOMIZED DESIGN

The GLM Procedure

Number of Observations Read	12
Number of Observations Used	12

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	35.71341269	7.14268254	6.79	0.0186
Error	6	6.30908731	1.05151455		
Corrected Total	11	42.02250000			

R-Square	Coeff Var	Root MSE	y Mean
0.849864	19.43950	1.025434	5.275000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	11.60121622	11.60121622	11.03	0.0160
B	1	12.62626263	12.62626263	12.01	0.0134
A*A	1	1.71667962	1.71667962	1.63	0.2486
B*B	1	5.35925423	5.35925423	5.10	0.0648
A*B	1	4.42000000	4.41000000	4.19	0.0865

Table 12.8 (Continued)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	11.60121622	11.60121622	11.03	0.0160
B	1	12.62626263	12.62626263	12.01	0.0134
A*A	1	5.54678825	5.54678825	5.28	0.0614
B*B	1	5.35925423	5.35925423	5.10	0.0648
A*B	1	4.41000000	4.41000000	4.19	0.0865

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	3.562529219	0.72492171	4.91	0.0027
A	0.989864865	0.29801066	3.32	0.0160
B	1.262626263	0.36437205	3.47	0.0134
A*A	1.010640582	0.44003131	2.30	0.0614
B*B	1.083796071	0.48006831	2.26	0.0648
A*B	1.050000000	0.51271692	2.05	0.0865

CENTRAL COMPOSITE DESIGN
AS SPLIT-PLOT DESIGN

3

The Mixed Procedure

Model Information

Data Set	WORK.CCD
Dependent Variable	y
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
block	6	1 2 3 4 5 6

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	29.29778783	
1	1	29.09016318	0.00000000

Convergence criteria met.

Cov Parm	Estimate
block	0.3065
Residual	0.8472

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	3.5625	0.8543	2	4.17	0.0530
A	0.9899	0.3512	2	2.82	0.1062
B	1.2626	0.3271	4	3.86	0.0181
A*A	1.0106	0.5185	2	1.95	0.1906
B*B	1.0838	0.5657	2	1.92	0.1955
A*B	1.0500	0.4602	4	2.28	0.0846

12.9 EXERCISES

12.1 Consider an experiment with 5 input variables.

- (i) Obtain an appropriate $1/2$ fraction of the 2^5 factorial to fit a first-order model.
- (ii) For the design chosen in (i) sketch the ANOVA table assuming that each design point is replicated twice.
- (iii) Suppose we need to run the experiment in blocks of size 8. Write out an appropriate plan, give the associated linear model and outline the ANOVA table.

12.2 Consider a simplex design with $k = 4$ input variables.

- (i) Write out explicitly the design-model matrix D .
- (ii) Outline the ANOVA table with $r = 2$ replications for each design point.

12.3 Consider a central composite design for an experiment with five input variables. Show that with a $1/2$ fraction of resolution V of the 2^5 as the factorial part of the design all linear, quadratic and linear \times linear effects [see model (12.15)] can be estimated.

12.4 For an experiment with four input variables, construct a Box-Behnken design using the following BIBD with four treatments and six blocks of size 2:

Treatment	Blocks					
	1	2	3	4	5	6
1	x	x	x			
2	x			x	x	
3		x		x		x
4			x		x	x

This Page Intentionally Left Blank

CHAPTER 13

Split-Plot Type Designs

13.1 INTRODUCTION

In all the error-control designs discussed so far we have had one type of EU for all the treatments and one randomization process to assign the treatments to the EUs. There exist, however, many situations where for a factorial experiment different types of EUs are being used and where the levels of some factors are applied sequentially, necessitating separate randomization procedures. We have already pointed to such situations in Sections 12.4.6 and 2.3.2. In the simplest case we have EUs of one size for the levels of one of two factors. Those EUs are then subdivided into smaller EUs to which the levels of the second factor are applied. This procedure is referred to as the *split-unit principle*. The following is an example of such a situation.

EXAMPLE 13.1: Suppose we want to investigate the breaking strength of dinnerware manufactured by using different chemical compounds and baking it at different temperatures. Let temperature be factor A with three levels, say $a_1 = 400^\circ$, $a_2 = 500^\circ$, $a_3 = 600^\circ$, and let factor C denote chemical compounds with levels c_1, c_2, c_3, c_4 , say, each being a specified chemical compound. We have three furnaces available. Each furnace will be set at one of the randomly assigned temperatures. In each furnace we then place four dinner plates each individually produced using a different (randomly assigned) chemical compound for each plate. This process is repeated on several days. For each plate the breaking strength is then determined using a suitable machine.

The large EUs are furnaces and the smaller EUs are the dinner plates:

	Furnace					
	a_3		a_1		a_2	
Day 1	c_3	c_4	c_1	c_4	c_3	c_1
	c_2	c_1	c_2	c_3	c_2	c_4
	a_1		a_3		a_2	
Day 2	c_4	c_1	c_2	c_1	c_4	c_2
	c_2	c_3	c_4	c_3	c_3	c_1
	etc.					

If we repeat this process on r days we have r replications of each temperature, but $3r$ replications for each chemical compound. \square

Not only will this type of arrangement lead to different precisions for the comparisons among the levels of the A -factor and among those of the C -factor, but the fact that the two factors are associated with different types of EUs leads to different experimental error variances associated with these comparisons. This is the reason why these types of experiments must be distinguished very carefully from the factorial experiments described in Chapter 11 (from a purely technical point of view there exists a link between these two types of experiments through the notion of interblock information, which is discussed in Chapters II.7 – 11).

We shall now describe some specific forms of designs which use different types of experimental units.

13.2 SIMPLE SPLIT-PLOT DESIGN

This design was developed and used first and foremost for agricultural, mainly agroeconomic experiments (see Yates, 1935 and 1937), but its applicability goes now across all fields of experimental research. Even so, the terminology for this design still makes references to plots of various types, but the reader should have no difficulty translating this into any other subject matter area.

13.2.1 Superimposing Two Randomized Complete Block Designs

We have two treatment factors A and B , with levels a_1, a_2, \dots, a_a and b_1, b_2, \dots, b_b , respectively. Factor A is referred to as the *whole-plot factor* and the EUs to which the levels of A are applied are the *whole-plots*. Factor B is the *split-plot factor* and the EUs to which the levels of B are applied are the *split-plots*, each whole-plot having b split-plots as illustrated below for $b = 4$:



A replicate consists then of one application of each level a_1, a_2, \dots, a_a and within each of the a whole-plots of one application of each level b_1, b_2, \dots, b_b . And the design consists then of r such replications.

It is useful to think of this arrangement as superimposing one RCBD on top of another RCBD. For the first RCBD, involving the whole-plots and the whole-plot factor, we have

$$\text{RCBD}_A : \quad t = a, \text{ number of blocks} = r$$

and for the second RCBD, involving the split-plots and the split-plot factor, we have

$$\text{RCBD}_B : \quad t = b, \text{ number of blocks} = ra$$

This brings out the fact that two independent randomizations are being used.

This structure suggests the partitioning of the $rab - 1$ d.f. available from the rab observations in the following way. If we consider first the $RCBD_B$ we have the partition

Source	d.f.
Blocks (whole plots)	$ra - 1$
B -factor	$b - 1$
Residual (B)	$(ra - 1)(b - 1)$
Total	$rab - 1$

We realize, however, that the systematic differences among blocks in a replicate are due only to the different levels of factor A . This and the fact that the replicates form the blocks for the $RCBD_A$ implies that we have the following partition of the $ra - 1$ d.f. for whole-plots

Source	d.f.
Replicates	$r - 1$
A -factor	$a - 1$
Error (A)	$(r - 1)(a - 1)$
Whole-plots	$ra - 1$

It follows from this partitioning that the $(ra - 1)(b - 1)$ d.f. for Residual (B) can be partitioned further into

Source	d.f.
Replicates $\times B$	$(r - 1)(b - 1)$
$A \times B$	$(a - 1)(b - 1)$
Error(A) $\times B$	$(r - 1)(a - 1)(b - 1)$
Residual (B)	$(ra - 1)(b - 1)$

Assuming no replicate $\times B$ interaction (since we are assuming unit-treatment additivity), we then have the complete partitioning of the d.f. as given in the ANOVA of Table 13.1. The associated sums of squares and their properties can be derived as follows, based on the observations arising from the rab split-plots.

Table 13.1 ANOVA for Split-Plot Design

Source	d.f.	SS	$E(\text{MS})$
Replicates	$r - 1$	$ab \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \text{SS}(R)$	
A-factor	$a - 1$	$rb \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = \text{SS}(A)$	$\sigma_{eB}^2 + b\sigma_{eA}^2 + rb \sum_j \alpha_j^2 / (a - 1)$
Error (A)	$(r - 1)(a - 1)$	$b \sum_{i,j} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 = \text{SS}(E_A)$	$\sigma_{eB}^2 + b\sigma_{eA}^2$
B-factor	$b - 1$	$ra \sum_k (\bar{y}_{.k.} - \bar{y}_{...})^2 = \text{SS}(B)$	$\sigma_{eB}^2 + ra \sum_k \beta_k^2 / (b - 1)$
$A \times B$	$(a - 1)(b - 1)$	$r \sum_{j,k} (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{.k.} + \bar{y}_{...})^2 = \text{SS}(A \times B)$	$\sigma_{eB}^2 + r \sum_{j,k} (\alpha\beta)_{jk}^2 / [(a - 1)(b - 1)]$
Error (B)	$(r - 1)a(b - 1)$	$\sum_{ij,k} (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{.jk} + \bar{y}_{.j.})^2 = \text{SS}(E_B)$	σ_{eB}^2
Total	$rab - 1$	$\sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2$	

13.2.2 Derived Linear Model

We denote the conceptual response of the v th split-plot in the u th whole-plot of the i th replicate to which the j th level of the whole-plot factor A and the k th level of the split-plot factor B have been applied by x_{iuvjk} . Assuming unit-treatment additivity we write

$$x_{iuvjk} = U_{iuv} + T_{jk}, \quad (13.1)$$

where U_{iuv} is the unit contribution and T_{jk} is the treatment contribution. We then write further, using obvious notation,

$$U_{iuv} = \bar{U}_{...} + (\bar{U}_{i..} - \bar{U}_{...}) + (\bar{U}_{iu.} - \bar{U}_{i..}) + (U_{iuv} - \bar{U}_{iu.}) \quad (13.2)$$

and

$$T_{jk} = \bar{T}_{..} + (\bar{T}_{j.} - \bar{T}_{..}) + (\bar{T}_{.k} - \bar{T}_{..}) + (T_{jk} - \bar{T}_{j.} - \bar{T}_{.k} + \bar{T}_{..}). \quad (13.3)$$

Substituting (13.2) and (13.3) into (13.1) and defining

$$\bar{U}_{...} + \bar{T}_{..} = \mu$$

$$\bar{U}_{i..} - \bar{U}_{...} = r_i$$

the effect of the i th replicate,

$$\bar{T}_{j.} - \bar{T}_{..} = \alpha_j$$

the effect of the j th level of A ,

$$\bar{T}_{.k} - \bar{T}_{..} = \beta_k$$

the effect of the k th level of B ,

$$T_{jk} - \bar{T}_{j.} - \bar{T}_{.k} + \bar{T}_{..} = (\alpha\beta)_{jk}$$

the interaction effect between the j th level of A and the k th level of B , we obtain

$$x_{iuvjk} = \mu + r_i + \alpha_j + (\bar{U}_{iu.} - \bar{U}_{i..}) + \beta_k + (\alpha\beta)_{jk} + (U_{iuv} - \bar{U}_{iu.}). \quad (13.4)$$

We actually observe y_{ijk} , the response of the (jk) treatment combination in replicate i . The observed and conceptual responses are linked to each other by two design random variables associated with the randomization processes of factors A and B , respectively. Let

$$\delta_{iu}^j = \begin{cases} 1 & \text{if level } j \text{ of factor } A \text{ is applied to the } u\text{th whole-plot in replicate } i \\ 0 & \text{otherwise} \end{cases}$$

and

$$\delta_{iuv}^{jk} = \begin{cases} 1 & \text{if level } k \text{ of factor } B \text{ is applied to the } v\text{th split-plot in the } u\text{th} \\ & \text{whole-plot of replicate } i \text{ given that the } j\text{th level of } A \text{ has been applied to} \\ & \text{that whole-plot} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$y_{ijk} = \sum_{u=1}^a \delta_{iu}^j \left(\sum_{v=1}^b \delta_{iuv}^{jk} x_{iuvjk} \right). \quad (13.5)$$

Substituting (13.4) into (13.5) we obtain

$$y_{ijk} = \mu + r_i + \alpha_j + \eta_{ij} + \beta_k + (\alpha\beta)_{jk} + \zeta_{ijk}, \quad (13.6)$$

where

$$\eta_{ij} = \sum_u \delta_{iu}^j (\bar{U}_{iu.} - \bar{U}_{i..})$$

and

$$\zeta_{ijk} = \sum_{u,v} \delta_{iu}^j \delta_{iuv}^{jk} (U_{iuv} - \bar{U}_{iu.}).$$

The η_{ij} and ζ_{ijk} are the two errors arising from the fact that we have two types of EUs and two independent randomization processes. Following arguments similar to those given for the RCBD (Chapter 9) we can derive easily the distributional properties of these errors. For example, it is obvious that

$$E_R(\eta_{ij}) = 0, \quad E_R(\zeta_{ijk}) = 0$$

and that the η_{ij} 's are correlated and the ζ_{ijk} 's are correlated. Both types of errors constitute only unit errors to which we may add the technical errors, in this case treatment errors for factors A and B , respectively, and observational error. We indicate this by rewriting (13.6) to obtain the final model

$$y_{ijk} = \mu + r_i + \alpha_j + e_{ij}^A + \beta_k + (\alpha\beta)_{jk} + e_{ijk}^B. \quad (13.7)$$

For all purposes of inference about the treatment effects we may treat the e_{ij}^A and e_{ijk}^B as if they were i.i.d. with means 0 and variances σ_{eA}^2 and σ_{eB}^2 , respectively. This leads to the $E(\text{MS})$ in Table 13.1.

13.2.3 Testing of Hypotheses

The forms of the $E(\text{MS})$ indicate the appropriate tests of significance. Relying on the approximation of the randomization test by the F -test we test

(i) $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$ by

$$F = \frac{\text{MS}(A)}{\text{MS}(E_A)} \sim F_{a-1, (r-1)(a-1)};$$

(ii) $H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$ by

$$F = \frac{\text{MS}(B)}{\text{MS}(E_B)} \sim F_{b-1, (r-1)a(b-1)};$$

(iii) $H_0: \text{all } (\alpha\beta)_{jk} = 0$ by

$$F = \frac{\text{MS}(A \times B)}{\text{MS}(E_B)} \sim F_{(a-1)(b-1), (r-1)a(b-1)}.$$

13.2.4 Estimating Treatment Contrasts

Since the split-plot design is an orthogonal design, treatment contrasts are estimated simply by the corresponding contrasts of appropriate treatment means and their variances are obtained by using model (13.7) and its properties:

- (i) The contrast $\sum_j c_j \alpha_j$ ($\sum_j c_j = 0$) among whole-plot treatment effects is estimated by $\sum_j c_j \bar{y}_{..j}$. Since

$$\text{var}(\bar{y}_{ij.}) = \frac{1}{b}(\sigma_{eB}^2 + b\sigma_{eA}^2)$$

for every $i = 1, 2, \dots, r; j = 1, 2, \dots, a$, we have

$$\text{var} \left(\sum_j c_j \bar{y}_{..j} \right) = \sum_j c_j^2 \frac{\sigma_{eB}^2 + b\sigma_{eA}^2}{rb} \quad (13.8)$$

and hence, from Table 13.1,

$$\hat{\text{var}} \left(\sum_j c_j \hat{\alpha}_j \right) = \sum_j c_j^2 \frac{\text{MS}(E_A)}{rb}. \quad (13.9)$$

- (ii) The contrast $\sum_k d_k \beta_k$ ($\sum_k d_k = 0$) among the split-plot treatment effects is estimated by $\sum_k d_k \bar{y}_{..k}$. Now

$$\begin{aligned} \sum_k d_k \bar{y}_{..k} &= \frac{1}{ra} \sum_k d_k \sum_{i,j} y_{ijk} \\ &= \frac{1}{ra} \sum_k d_k \left[ra\mu + \sum_{i,j} e_{ij}^A + ra\beta_k + \sum_{i,j} e_{ijk}^B \right] \\ &= \sum_k d_k \beta_k + \frac{1}{ra} \sum_{i,j,k} d_k e_{ijk}^B \end{aligned}$$

and hence

$$\text{var} \left(\sum_k d_k \bar{y}_{..k} \right) = \frac{1}{ra} \sum_k d_k^2 \sigma_{eB}^2 \quad (13.10)$$

and

$$\hat{\text{var}} \left(\sum_k d_k \hat{\beta}_k \right) = \sum_k d_k^2 \frac{\text{MS}(E_B)}{ra}. \quad (13.11)$$

- (iii) We now consider a contrast among split-plot effects averaged over a set of p ($p < a$) whole-plot treatments. If we write

$$\tau_{jk} = \alpha_j + \beta_k + (\alpha\beta)_{jk} \quad (13.12)$$

and

$$\beta_k^{(p)} = \frac{1}{p} \sum_{j=1}^p \tau_{jk}$$

(to simplify notation we have averaged over the first p whole-plot treatments), we then consider $\sum_k d_k \beta_k^{(p)}$ with $\sum d_k = 0$. The estimator for this contrast is

$$\begin{aligned} \frac{1}{p} \sum_k d_k \left[\sum_{j=1}^p \bar{y}_{jk} \right] &= \frac{1}{rp} \sum_k d_k \left[\sum_{i,j} y_{ijk} \right] \\ &= \frac{1}{rp} \sum_k d_k \left[rp\mu + r \sum_{j=1}^p \alpha_j + \sum_{i,j} e_{ij}^A + rp\beta_k + r \sum_{j=1}^p (\alpha\beta)_{jk} + \sum_{i,j} e_{ijk}^B \right] \\ &= \sum_k d_k \left[\beta_k + \frac{1}{p} \sum_j (\alpha\beta)_{jk} \right] + \frac{1}{rp} \sum_{i,j,k} d_k e_{ijk}^B \end{aligned}$$

with

$$\begin{aligned} E \left(\sum_k d_k \hat{\beta}_k^{(p)} \right) &= \sum_k d_k \left[\beta_k + \frac{1}{p} \sum_j (\alpha\beta)_{jk} \right] \\ \text{var} \left(\sum_k d_k \hat{\beta}_k^{(p)} \right) &= \frac{1}{rp} \sum_k d_k^2 \sigma_{eB}^2 \end{aligned} \quad (13.13)$$

and

$$\text{var} \left(\sum_k d_k \hat{\beta}_k^{(p)} \right) = \sum_k d_k^2 \frac{\text{MS}(E_B)}{rp}. \quad (13.14)$$

- (iv) We may also consider a contrast among whole-plot treatment effects averaged over a set of q ($q < b$) split-plot treatments. If we write, using (13.12),

$$\alpha_j^{(q)} = \frac{1}{q} \sum_{k=1}^q \tau_{jk}$$

(to simplify notation we have averaged over the first q split-plot treatments), we

are then interested in $\sum_j c_j \alpha_j^{(q)}$ with $\sum_j c_j = 0$. The estimator for this contrast is

$$\begin{aligned} \frac{1}{q} \sum_j c_j \left[\sum_{k=1}^q \bar{y}_{jk} \right] &= \frac{1}{rq} \sum_j c_j \left[\sum_{i,k} y_{ijk} \right] \\ &= \frac{1}{rq} \sum_j c_j \left[rq\mu + rq\alpha_j + q \sum_i e_{ij}^A + r \sum_{k=1}^q \beta_k + r \sum_{k=1}^q (\alpha\beta)_{jk} + \sum_{i,k} e_{ijk}^B \right] \\ &= \sum_j c_j \alpha_j + \frac{1}{r} \sum_{i,j} c_j e_{ij}^A + \frac{1}{q} \sum_{j,k} c_j (\alpha\beta)_{jk} + \frac{1}{rq} \sum_{i,j,k} c_j e_{ijk}^B \end{aligned}$$

with

$$\begin{aligned} E \left(\sum_j c_j \hat{\alpha}_j^{(q)} \right) &= \sum_j c_j^2 \left[\alpha_j + \frac{1}{q} \sum_k (\alpha\beta)_{jk} \right] \\ \text{var} \left(\sum_j c_j \hat{\alpha}_j^{(q)} \right) &= \frac{1}{r} \sum_j c_j^2 \sigma_{eA}^2 + \frac{1}{rq} \sum_j c_j^2 \sigma_{eB}^2. \end{aligned} \quad (13.15)$$

We know from Table 13.1 that

$$\hat{\sigma}_{eB}^2 = \text{MS}(E_B) \quad (13.16)$$

and

$$\hat{\sigma}_{eA}^2 = \frac{1}{b} [\text{MS}(E_A) - \text{MS}(E_B)] \quad (13.17)$$

(if $\text{MS}(E_B) > \text{MS}(E_A)$ we take $\hat{\sigma}_{eA}^2 = 0$).

We then estimate (13.15) as

$$\begin{aligned} \text{var} \left(\sum_j c_j \hat{\alpha}_j^{(q)} \right) &= \frac{1}{rq} \sum_j c_j^2 [q \hat{\sigma}_{eA}^2 + \hat{\sigma}_{eB}^2] \\ &= \frac{1}{rq} \sum_j c_j^2 \left[q \frac{\text{MS}(E_A) - \text{MS}(E_B)}{b} + \text{MS}(E_B) \right] \\ &= \frac{1}{rb} \sum_j c_j^2 \left[\text{MS}(E_A) + \frac{b-q}{q} \text{MS}(E_B) \right]. \end{aligned} \quad (13.18)$$

(v) Occasionally, a contrast of the form $\tau_{jk} - \tau_{j'k'} (j \neq j')$ is of interest. Obviously,

$$\begin{aligned} \hat{\tau}_{jk} - \hat{\tau}_{j'k'} &= \bar{y}_{jk} - \bar{y}_{j'k'} \\ &= \alpha_j - \alpha_{j'} + \frac{1}{r} \sum_i (e_{ij}^A - e_{ij'}^A) + \beta_k - \beta_{k'} \\ &\quad + (\alpha\beta)_{jk} - (\alpha\beta)_{j'k'} + \frac{1}{r} \sum_i (e_{ijk}^B - e_{ij'k'}^B) \end{aligned}$$

with

$$\text{var}(\hat{\tau}_{jk} - \hat{\tau}_{j'k'}) = \frac{2}{r}\sigma_{eA}^2 + \frac{2}{r}\sigma_{eB}^2 \quad (13.19)$$

and, using (13.16) and (13.17),

$$\text{var}(\hat{\tau}_{jk} - \hat{\tau}_{j'k'}) = \frac{2}{rb}[\text{MS}(E_A) + (b-1)\text{MS}(E_B)]. \quad (13.20)$$

13.2.5 Testing Hypotheses about Treatment Contrasts

Of the contrasts described above, (i) and (ii) are usually of most interest. Tests of significance about or confidence intervals for them can be obtained by referring to t -statistics with the appropriate d.f., $(r-1)(a-1)$ for (i) and $(r-1)a(b-1)$ for (ii). We should mention, however, that inferences about the α_j 's or the β_k 's may not always be meaningful when the $A \times B$ interaction is significant. Careful examination of the kind of interaction present is necessary before proceeding to the inference about main effects. If such inferences are not appropriate contrasts of the form described in (iii) and (iv) may be more useful, often with $p = q = 1$. There is no difficulty in dealing with (iii) using the t -statistic with $(r-1)a(b-1)$ d.f. However, there does not exist an exact test for the contrast given in (iv). A reasonable method to use is to form the t -statistic in the usual way, that is,

$$t = \frac{\sum_j c_j \hat{\alpha}_j^{(q)}}{\left[\text{var} \left(\sum_j c_j \hat{\alpha}_j^{(q)} \right) \right]^{1/2}} \quad (13.21)$$

and then compare (13.21) with the following weighted critical t -value:

$$t_\alpha = \frac{\text{MS}(E_A)t_{(r-1)(a-1),\alpha} + \frac{b-q}{q}\text{MS}(E_B)t_{(r-1)a(b-1),\alpha}}{\text{MS}(E_A) + \frac{b-q}{q}\text{MS}(E_B)}, \quad (13.22)$$

where $t_{\nu,\alpha}$ refers to the α -percentage point of the t -distribution with ν d.f. (for example, Cochran and Cox, 1957).

Another method to use is that suggested by Satterthwaite (1947), that is, to compute (13.21) and approximate its distribution by that of a t -statistic with ν d.f. where

$$\nu = \frac{\left[\text{MS}(E_A) + \frac{b-q}{q}\text{MS}(E_B) \right]^2}{\frac{[\text{MS}(E_A)]^2}{(r-1)(a-1)} + \frac{\left[\frac{b-q}{q}\text{MS}(E_B) \right]^2}{(r-1)a(b-1)}}. \quad (13.23)$$

The same procedure can be used for inference about a contrast of type (v), using $q = 1$. As mentioned earlier, this comparison is only occasionally of interest, for example to compare a control treatment with a particular treatment combination.

We conclude this section by pointing out that the whole-plot treatments themselves can have a factorial structure and the same is possible for the split-plot treatments. The reader should have no difficulty modifying the ANOVA and hence making use of the factorial structures to analyze such an experiment.

13.3 RELATIVE EFFICIENCY OF SPLIT-PLOT DESIGN

Under most circumstances the split-plot design is used for purely technical and practical reasons, as the levels of some factor can be applied only to large EUs which can then be “split” into smaller EUs for application of the levels of the other factor. This includes also the distinction between hard-to-change and easy-to-change factors in industrial experimentation (see Section 12.4.6). It is, however, of interest to evaluate the efficiency of the split-plot design relative to the RCBD with r blocks. The question then is: Given that we have carried out a split-plot experiment, what would have been $MS(E)$ for the RCBD? This, of course, determines how much information would have been available for all treatment comparisons. We see from Table 13.2, using a uniformity trial for both situations, that is, pooling treatment sums of squares with appropriate error sums of squares, that

$$r(ab - 1)MS(E) = r(a - 1)MS(E_A) + ra(b - 1)MS(E_B)$$

and hence

$$MS(E) = \frac{(a - 1)MS(E_A) + a(b - 1)MS(E_B)}{ab - 1}. \quad (13.24)$$

The information on all treatment comparisons in a RCBD would then have been proportional to $1/MS(E)$, whereas in the split-plot design information on whole-plot treatment comparisons is proportional to $1/MS(E_A)$ and on split-plot treatment comparisons and interaction proportional to $1/MS(E_B)$. Since $MS(E)$ is a weighted average of $MS(E_A)$ and $MS(E_B)$, and since $MS(E_A)$ is usually greater than $MS(E_B)$ (except for sampling errors), $MS(E)$ will be intermediate in size between $MS(E_A)$ and $MS(E_B)$.

We can then state the results concerning relative efficiencies of the split-plot design versus the RCBD as follows: For A -factor comparisons we have

$$ERE_A(\text{Split-plot design vs. RCBD}) = \frac{MS(E)}{MS(E_A)} < 1 \quad (13.25)$$

and for B -factor and $A \times B$ comparisons we have

Table 13.2 ANOVA for Uniformity Trial

(a) Split-plot design		
Source	d.f.	MS
Replicates	$r - 1$	
Error (A)	$r(a - 1)$	$MS(E_A)$
Error (B)	$ra(b - 1)$	$MS(E_B)$
Total	$rab - 1$	
(b) RCBD		
Source	d.f.	MS
Blocks	$r - 1$	
Error	$r(ab - 1)$	$MS(E)$
Total	$rab - 1$	

$$ERE_B(\text{Split-plot design vs. RCBD}) = \frac{MS(E)}{MS(E_B)} > 1. \quad (13.26)$$

Results (13.25) and (13.26) express the obvious: Although the average information is the same for both designs, the information on whole-plot treatment comparisons is less precise in the split-plot design than in the RCBD, whereas the opposite is true for split-plot treatment and interaction comparisons. Hence, unless practical reasons dictate the use of a split-plot design or one is more interested in one factor than the other, use of a RCBD seems preferable.

13.4 OTHER FORMS OF SPLIT-PLOT DESIGNS

We mentioned in Section 13.2 that in order to better understand the structure of the simple split-plot design it is advantageous to view it as superimposing one RCBD (for the split-plot treatments) on top of another RCBD (for the whole-plot treatments). We shall refer to this as a SPD(RCBD, RCBD). Variations of this form of split-plot design are possible by using different component designs, other than both RCBD. Some useful combinations are indicated below (where IBD refers to incomplete block design) and discussed in the section indicated.

Error-control design for whole-plot treatment	Error-control design for split-plot treatment	Section
CRD	RCBD	13.4.1
CRD	LSD	13.4.3
LSD	RCBD	13.4.4
CRD	IBD	13.4.5
GRBD	RCBD	13.4.6
GRBD	IBD	13.4.7
IBD	RCBD	13.4.8
RCBD	GRBD	13.4.9

13.4.1 SPD(CRD, RCBD)

Each level of the A -factor is randomly assigned to r whole-plots and within each whole-plot the b levels of the B -factor are randomly applied to the split-plots. In this situation the whole-plots are often subjects. Each subject is given a certain treatment, that is, one of the levels of the whole-plot factor (A -factor) such that each level of the A -factor is applied at random to r subjects. Then each subject will receive, in random order, sequentially all b levels of the B -factor. It is for this reason that this type of design is often referred to as a *between-and-within-subjects design*, where the A -factor is referred to as the *between-subjects factor* and the B -factor is referred to as the *within-subjects factor*. A diagram of the structure of the data from such an experiment is given in Figure 13.1 (ignoring randomization).

A suitable model is of the form

$$y_{ijk} = \mu + \alpha_i + e_{ij}^A + \beta_k + (\alpha\beta)_{ik} + e_{ijk}^B \quad (13.27)$$

or

$$y_{ijk} = \mu + \alpha_i + s_{ij} + \beta_k + (\alpha\beta)_{ik} + e_{ijk}^B,$$

where s_{ij} represents the effect of the j th subject receiving the i th level of the A -factor, ($i = 1, 2, \dots, a$; $j = 1, 2, \dots, r$; $k = 1, 2, \dots, b$). The ANOVA is given in Table 13.3.

13.4.2 Split-Plot Design in Time

In some types of experiments subjects (EUs) are given a certain treatment, a dietary regimen for example. Observations (say weight) are then made at specified times (for instance, every month for one year). The design for such an experiment is usually of the form of an SPD(RCBD, RCBD) or SPD(CRD, RCBD) and is, therefore, often referred to as a “split-plot design in time,” where the treatment is the A -factor and the times are considered to be the “levels of the B -factor.” There are a number of problems with this viewpoint. First, the “ B -levels” are obviously not randomized. Secondly, and even more importantly, there exists a covariance structure for the observations and hence for the errors other than the one ordinarily induced by the randomization procedure. This may invalidate the analysis outlined above, and only if the covariance structure satisfies the Huynh-Feldt conditions (Huynh and Feldt, 1970) do $MS(B)/MS(E_B)$

A-factor		B-factor			
		1	2	...	b
1	S_{11}	y_{111}	y_{112}		y_{11b}
	S_{12}	y_{121}	y_{122}		y_{12b}
	\vdots				
	S_{1r}	y_{1r1}	y_{1r2}		y_{1rb}
2	S_{21}	y_{211}	y_{212}		y_{21b}
	S_{22}	y_{221}	y_{222}		y_{22b}
	\vdots				
	S_{2r}	y_{2r1}	y_{2r2}		y_{2rb}
\vdots	\vdots				
a	S_{a1}	y_{a11}	y_{a12}		y_{a1b}
	S_{a2}	y_{a21}	y_{a22}		y_{a2b}
	\vdots				
	S_{ar}	y_{ar1}	y_{ar2}		y_{arb}

Figure 13.1 Between-and-within-subjects design.

Table 13.3 ANOVA for SPD(CRD, RCBD)

Source	d.f.	SS	E(MS)
A-factor	$a - 1$	$rb \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$\sigma_{eB}^2 + b\sigma_{eA}^2 + rb \sum \alpha_i^2 / (a - 1)$
Error (A)	$a(r - 1)$	$b \sum_{i,j} (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$\sigma_{eB}^2 + b\sigma_{eA}^2$
B-factor	$b - 1$	$ra \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$	$\sigma_{eB}^2 + ra \sum \beta_k^2 / (b - 1)$
$A \times B$	$(a - 1)(b - 1)$	$r \sum_{i,k} (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$	$\sigma_{eB}^2 + r \sum (\alpha\beta)_{ikj}^2 / [(a - 1) \cdot (b - 1)]$
Error (B)	$a(r - 1)(b - 1)$	$\sum_{ijk} (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} + \bar{y}_{i..})^2$	σ_{eB}^2
Total	$rab - 1$	$\sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2$	

Factor A		Order			
		1	2	3	4
a_1	S_{11}	b_1	b_2	b_4	b_3
	S_{12}	b_3	b_4	b_2	b_1
	S_{13}	b_4	b_1	b_3	b_2
	S_{14}	b_2	b_3	b_1	b_4
	S_{21}	b_4	b_1	b_3	b_2
a_2	S_{22}	b_1	b_2	b_4	b_3
	S_{23}	b_3	b_4	b_2	b_1
	S_{24}	b_2	b_3	b_1	b_4

Figure 13.2 Schematic representation of a SPD(CRD, LSD) with $a = 2$, $b = 4$.

and $MS(A \times B)/MS(E_B)$ have F -distributions. These designs are called *repeated measures designs* (for a discussion see Chapter 14).

13.4.3 SPD(CRD, LSD)

Experiments in psychology and human factors engineering are often performed using this kind of design or variations of it. Rather than assign the levels of the split-plot factor randomly to the split-plots within each whole-plot as in the SPD(CRD, RCBD), they are assigned according to a Latin square design as follows. First, each level of the A -factor is randomly assigned to $r = b$ whole-plots. If it is suspected that the order of application of the B -levels within each whole-plot has a systematic effect on the outcome then a Latin square arrangement of the following type may be used. For each A -level the b whole-plots form the rows of an LSD and the orders of application form the columns of the LSD. For each A -level we thus have an LSD of size b and the b^2 row-column combinations represent the split-plots to which the B -levels are assigned according to a randomly selected $b \times b$ LSD. For $a = 2$ and $b = 4$ the design can be represented as in Figure 13.2.

The LSDs given in Figure 13.2 are actually of a specific type. They are sometimes referred to as *completely counter-balanced* or *diagram-balanced* (Wagenaar, 1969) and are constructed following a method due to Williams (1949) (see Section 10.7.2). The special feature of such an LSD is that each treatment precedes and follows every other treatment exactly once in the order of application. This is useful when learning effects or carry-over effects are suspected.

An example of the design described here might be a psychological experiment in which subjects (S_{ij}) are given different types of training (education), represented by the A -levels, and following that each subject performs sequentially a number of tasks (tests), the same for each subject and represented by the B -levels. It is suspected that a learning effect takes place. This means that subjects may respond differently to the

Table 13.4 ANOVA for SPD(CRD, LSD)

Source	d.f.	SS	E(MS)
<i>A</i> -factor	$a - 1$	$b^2 \sum_i (\bar{y}_{i...} - \bar{y}_{...})^2$	$\sigma_{eB}^2 + b\sigma_{eA}^2 + b^2 \sum_i \alpha_i^2 / (a - 1)$
Error (<i>A</i>)	$a(b - 1)$	$b \sum_{i,j} (\bar{y}_{ij..} - \bar{y}_{i...})^2$	$\sigma_{eB}^2 + b\sigma_{eA}^2$
Order (<i>O</i>)	$b - 1$	$ab \sum_k (\bar{y}_{...k} - \bar{y}_{...})^2$	
<i>B</i> -factor	$b - 1$	$ab \sum_l (\bar{y}_{...l} - \bar{y}_{...})^2$	$\sigma_{eB}^2 + ab \sum_l \beta_l^2 / (b - 1)$
$A \times B$	$(a - 1)(b - 1)$	$b \sum_{i,l} (\bar{y}_{i..l} - \bar{y}_{i...} - \bar{y}_{...l} + \bar{y}_{...})^2$	$\sigma_{eB}^2 + b \sum_{i,l} (\alpha\beta)_{il}^2 / (a - 1)(b - 1)$
$A \times O$	$(a - 1)(b - 1)$	$b \sum_{i,k} (\bar{y}_{ij..k} - \bar{y}_{i...} - \bar{y}_{...k} + \bar{y}_{...})^2$	$\sigma_{eB}^2 + b \sum_{i,k} (\alpha\gamma)_{ik}^2 / (a - 1)(b - 1)$
Error (<i>B</i>)	$a(b - 1)(b - 2)$	Difference	σ_{eB}^2
Total	$ab^2 - 1$	$\sum_{i,j,k(l)} (y_{ijk(l)} - \bar{y}_{...})^2$	

same task given at different times (order).

An appropriate model for this design can be written as

$$y_{ijk(l)} = \mu + \alpha_i + e_{ij}^A + \gamma_k + \beta_l + (\alpha\beta)_{il} + (\alpha\gamma)_{ik} + e_{ijk(l)}^B, \quad (13.28)$$

where γ_k represents the k th order effect ($i = 1, 2, \dots, a; j, k, l = 1, 2, \dots, b$). This leads to the ANOVA given in Table 13.4. Model (13.28) includes a term for factor $A \times$ order interaction, $(\alpha\gamma)_{ik}$. It reflects differences among the “learning curves” for the different levels of the A -factor. If such differences are assumed to not exist then $SS(A \times O)$ in Table 13.4 can be pooled with $SS(E_B)$.

Just as the SPD(CRD, RCBD) the SPD(CRD, LSD) is sometimes also referred to as a *between-and-within-subjects design* or *mixed factorial design* (for example, Keppel and Zedeck, 1989), where the A -factor is the between-subjects factor and the B -factor is the within-subjects factor.

The SPD(CRD, LSD) bears a certain resemblance to the replicated LSDs except that we have here two different randomization procedures. As explained for the SPD(RCBD, RCBD) this leads to two error terms rather than one for the repeated LSD (see Section 10.3).

13.4.4 SPD(LSD, RCBD)

The whole-plots are arranged in an $a \times a$ Latin square and the levels of the A -factor are assigned in accordance with the randomization procedures for the LSD (see Chapter 10). In each whole-plot the levels of the B -factor are applied to the split-plots according to a RCBD. A suitable model is of the form

$$y_{ijkl} = \mu + \rho_i + \gamma_j + \alpha_k + e_{ijk}^A + \beta_l + (\alpha\beta)_{kl} + e_{ijkl}^B \quad (13.29)$$

Table 13.5 ANOVA for SPD(LSD, RCBD)

Source	d.f.	SS	E(MS)
Rows	$a - 1$	$ab \sum_i (\bar{y}_{i...} - \bar{y}_{....})^2$	
Columns	$a - 1$	$ab \sum_j (\bar{y}_{.j..} - \bar{y}_{....})^2$	
A -factor	$a - 1$	$ab \sum_k (\bar{y}_{..k.} - \bar{y}_{....})^2$	$\sigma_{eB}^2 + b\sigma_{eA}^2 - ab \sum_k \alpha_k^2 / (a - 1)$
Error (A)	$(a - 1)(a - 2)$	$b \sum_{ijk} (\bar{y}_{ijk.} - \bar{y}_{i...} - \bar{y}_{.j..} - \bar{y}_{..k.} + 2\bar{y}_{....})^2$	$\sigma_{eB}^2 + b\sigma_{eA}^2$
B -factor	$b - 1$	$a^2 \sum_l (\bar{y}_{...l} - \bar{y}_{....})^2$	$\sigma_{eB}^2 + a^2 \sum \beta_l^2 / (b - 1)$
$A \times B$	$(a - 1)(b - 1)$	$a \sum_{k,l} (\bar{y}_{..kl} - \bar{y}_{..k.} - \bar{y}_{...l} + \bar{y}_{....})^2$	$\sigma_{eB}^2 + a \sum_{k,l} (\alpha\beta)_{kl}^2 / (a - 1)(b - 1)$
Error (B)	$a(a - 1)(b - 1)$	Difference	σ_{eB}^2
Total	$a^2b - 1$	$\sum_{ijkl} (y_{ijkl} - \bar{y}_{....})^2$	

($i, j, k = 1, 2, \dots, a; l = 1, 2, \dots, b$). The ANOVA is given in Table 13.5. Where applicable this is rather effective in increasing the precision for whole-plot treatment comparisons (Yates, 1935).

As an example for this design we can envision an agronomic experiment where the experimental material (for example, field) requires blocking in two directions (rows and columns). The A -factor may represent different soil treatments such as no-till, shallow plowing, deep plowing ($a = 3$). The B -factor could be different types of fertilizer. The layout for $b = 2$ is given in Figure 13.3.

13.4.5 SPD(CRD, IBD)

This design is useful if the number of split-plots in a whole-plot is less than b , say K . A suitable arrangement might then be that the r replications for each whole-plot treatment form an IBD for the split-plot treatments, the IBD (apart from randomization) being the same for each level of the A -factor. For example, Robinson (1967) considers the specific case of a BIBD ($b, r, K, R; \lambda$), that is, each split-plot treatment occurs $R(< r)$ times with each whole-plot treatment, and each pair of split-plot treatments occurs together λ times with each whole-plot treatment.

As an example consider $a = 3, b = 4$. And suppose each whole-plot contains only $K = 2$ split-plots. Suppose further that each level of the A -factor is replicated $r = 6$ times. We can then use the following BIBD (4, 6, 2, 3; 1) for the B -factor:

1	1	1	2	2	3
2	3	4	3	4	4

(where each column represents a block) in such a way that the two treatments in a block are assigned randomly to the split-plots in a whole-plot and for each level of the A -factor the six blocks are assigned at random to the six replicates. The layout (apart from randomization) is given in Figure 13.4.

The model for this design is the same as (13.27) except that not all combinations (ijk) occur, but the analysis becomes now more complicated since the A -factor and

Rows	Columns					
	1		2		3	
1	:		:		:	
	a_2		a_1		a_3	
	:		:		:	
	b_2	:	b_1	b_2	:	b_1
	:		:		:	
2	:		:		:	
	a_1		a_3		a_2	
	:		:		:	
	b_1	:	b_2	b_2	:	b_1
	:		:		:	
3	:		:		:	
	a_3		a_2		a_1	
	:		:		:	
	b_1	:	b_2	b_1	:	b_2
	:		:		:	

Figure 13.3 Layout for SPD(LSD, RCBD) for $a = 3$, $b = 2$.

the B -factor are no longer orthogonal to each other. Partial sums of squares have to be obtained using the methods of Chapter 9 and Chapter II.1. We shall not go into the details here, but refer to Robinson (1967). A sketch of the ANOVA is given in Table 13.6.

We mention here that the IBD as the split-plot design needs to be chosen carefully. By that we mean that each level of the B -factor must occur with each level of the A -factor. Otherwise we cannot estimate all the interaction terms, and hence the d.f. for $A \times B$ will be less than $(a - 1)(b - 1)$.

13.4.6 SPD(GRBD, RCBD)

This design is similar to the SPD(CRD, RCBD) with r' replications of each of the whole plot factor levels in that each replicate constitutes a SPD(CRD, RCBD). The main advantage of this design is that we can now test for $\text{Rep} \times A$, $\text{Rep} \times B$ and $\text{Rep} \times A \times B$ interaction, using the following model

$$\begin{aligned}
 y_{ijkl} = & \mu + r_i + \alpha_j + (r\alpha)_{ij} + e_{ijk}^A \\
 & + \beta_l + (\alpha\beta)_{jl} + (r\beta)_{il} + (r\alpha\beta)_{ijl} + e_{ijkl}^B
 \end{aligned} \tag{13.30}$$

($i = 1, 2, \dots, r$; $j = 1, 2, \dots, a$; $k = 1, 2, \dots, r'$; $l = 1, 2, \dots, b$). This is, of course,

Factor A		Factor B	
a_1	S_{11}	b_1	b_2
	S_{12}	b_1	b_3
	S_{13}	b_1	b_4
	S_{14}	b_2	b_3
	S_{15}	b_2	b_4
	S_{16}	b_3	b_4
a_2	S_{21}	b_1	b_2
	S_{22}	b_1	b_3
	S_{23}	b_1	b_4
	S_{24}	b_2	b_3
	S_{25}	b_2	b_4
	S_{26}	b_3	b_4
a_3	S_{31}	b_1	b_2
	S_{32}	b_1	b_3
	S_{33}	b_1	b_4
	S_{34}	b_2	b_3
	S_{35}	b_2	b_4
	S_{36}	b_3	b_4

Figure 13.4 Layout of SPD(CRD, IBD).

Table 13.6 Outline of ANOVA for
SPD(CRD, BIBD)

Source	d.f.
A -factor	$a - 1$
Error (A)	$a(r - 1)$
B -factor	$b - 1$
$A \times B$	$(a - 1)(b - 1)$
Error(B)	$a(Rb - b - r + 1)$
Total	$abR - 1$

Table 13.7 Outline of ANOVA for SPD(GRBD, RCBD)

Source	d.f.
Rep	$r - 1$
A	$a - 1$
$\text{Rep} \times A$	$(a - 1)(r - 1)$
Error(A)	$ar(r' - 1)$
B	$b - 1$
$A \times B$	$(a - 1)(b - 1)$
$\text{Rep} \times B$	$(r - 1)(b - 1)$
$\text{Rep} \times A \times B$	$(r - 1)(a - 1)(b - 1)$
Error(B)	$ar(r' - 1)(b - 1)$
Total	$rr'ab - 1$

important if the replication factor is an intrinsic factor and if the type of interactions mentioned above are important. The outline of the ANOVA is given in Table 13.7.

13.4.7 SPD(GRBD, IBD)

This design is similar to the SPD(CRD, IBD) in that each replicate constitutes a SPD (CRD, IBD). In each replicate each whole-plot treatment is applied to, say, r' whole-plots and superimposed upon these is then an IBD, for example, a BIBD or PBIBD, for the split-plot treatments. This, too, is a nonorthogonal design and sums of squares in the ANOVA must be obtained from first principles (see Chapter 4 and also Chapter II.1). An outline of the ANOVA is given in Table 13.8 for the SPD(GRBD, BIBD($b, r', K, R; \lambda$)).

A special application of a SPD(GRBD, IBD) arises, for example, when the B -factor (that is, the split-plot factor) itself has a factorial structure and a system of confounding has to be used. To illustrate such a procedure we give a simple example.

Suppose we have three levels a_1, a_2, a_3 for the A -factor and a 2^3 factorial for the B -factor. Let us denote those factors by C with levels c_0, c_1 , D with levels d_0, d_1 , and E with levels e_0, e_1 . Suppose now we have $r' = 2$ applications of each level of A in each replicate and we have whole-plots with only four split-plots. Using the methods discussed in Chapter 11 the procedure to use is quite straightforward, namely to confound the 3-factor interaction CDE with whole-plots, assuming that this interaction is of less importance than main effects and 2-factor interactions. This leads (apart from randomization) to the following arrangement for one replicate:

a_1	a_1	a_2	a_2	a_3	a_3
$c_1d_0e_0$	$c_0d_0e_0$	$c_1d_0e_0$	$c_0d_0e_0$	$c_1d_0e_0$	$c_0d_0e_0$
$c_0d_1e_0$	$c_1d_1e_0$	$c_0d_1e_0$	$c_1d_1e_0$	$c_0d_1e_0$	$c_1d_1e_0$
$c_0d_0e_1$	$c_1d_0e_1$	$c_0d_0e_1$	$c_1d_0e_1$	$c_0d_0e_1$	$c_1d_0e_1$
$c_1d_1e_1$	$c_0d_1e_1$	$c_1d_1e_1$	$c_0d_1e_1$	$c_1d_1e_1$	$c_0d_1e_1$

**Table 13.8 Outline of ANOVA for
SPD(GRBD, BIBD)**

Source	d.f.
Replicates	$r - 1$
A -factor	$a - 1$
Error (A)	$(a - 1)(r - 1) + ar(r' - 1)$
B -factor	$b - 1$
$A \times B$	$(a - 1)(b - 1)$
Error(B)	$a(rRb - b - rr' + 1)$
Total	$arbR - 1$

The ANOVA for this design is given in Table 13.9. This is an orthogonal design and hence all sums of squares are easily obtainable using the usual procedures. The important feature of this design is that the five d.f. among whole-plots within a replicate can be partitioned into d.f. for A , CDE (since it is confounded with whole plots), and $A \times$ CDE. This is an example of “recovery of interblock information,” a procedure discussed in Chapters II.1 and 8–11.

13.4.8 SPD(IBD, RCBD)

This situation may arise in the following context: Suppose we want to investigate and compare different therapeutic treatments consisting of a combination of inoculation and ointment. We propose to use identical twins for this study. We have three different substances, say a_1, a_2, a_3 , for the inoculation, each individual receiving one substance, and we have two ointments, say b_1, b_2 , each being applied to one arm of each individual. In other words, the whole-plots are the individuals and the split-plots are the arms of each individual. Schematically the arrangement of the treatment combinations may be represented as follows:

Twin pair	Individual	Innoculate	Arm	
			Left	Right
1	1	a_2	b_1	b_2
	2	a_1	b_2	b_1
2	1	a_1	b_2	b_1
	2	a_3	b_2	b_1
3	1	a_3	b_1	b_2
	2	a_2	b_2	b_1

This basic pattern may be replicated r times, using proper randomization. The IBD used here is obviously a BIBD $(3, 3, 2, 1; 1)$ or, for the entire experiment, a BIBD $(3, 3r, 2, 2r; r)$. The ANOVA for this design is outlined in Table 13.10. Again, this is a nonorthogonal design.

Table 13.9 ANOVA for SPD(GRBD, IBD),
Using a System of Confounding

Source	d.f.	$E(MS)$
Replicates	$r - 1$	
A	2	
CDE	1	
$A \times CDE$	2	
Error (A)	$5(r - 1)$	$\sigma_{eB}^2 + r\sigma_{eA}^2$
C, D, E	6	
CD, CE, DE		
$A \times C, A \times D, A \times E,$		
$A \times CD, A \times CE,$	12	
$A \times DE$		
Error(B)	$18(r - 1)$	σ_{eB}^2
Total	$24r - 1$	

13.4.9 SPD(RCBD, GRBD)

If the whole-plots can be divided into more than B sub-plots then the design for the split-plot factor B may be a GRBD with r' replicates for each of its b levels, or some form of extended block design as discussed in Section 9.8.5

Table 13.10 Outline of ANOVA for SPD(BIBD, RCBD) using BIBD ($3, 3r, 2, 2r; r$)

Source	d.f.
Replicates	$r - 1$
Pairs/replicates	$2r$
A -factor	2
Error (A)	$6r - 3 - 3r + 1 = 3r - 2$
B -factor	1
$A \times B$	2
Error(B)	$(6r - 1) - 2 = 3(2r - 1)$
Total	$12r - 1$

An outline of the ANOVA for the SPD(RCBD, GRBD) is given in Table 13.11. We have included here $\text{Rep} \times B$ and $\text{Rep} \times A \times B$ as separate sources of variation. They may, of course, be pooled with the Error(B) if these interactions are considered to be negligible.

13.4.10 Summary

The designs given in this section represent obviously only a few examples of different forms of split-plot designs. The reader should have no difficulty thinking of other examples or of considering the examples given above more generally. The important point is that it is useful to represent split-plot designs as superimposing two suitable error-reduction designs. Those component designs should be chosen to best suit the experimental situation present.

13.5 SPLIT-BLOCK DESIGN

Unfortunately, the terminology for error-reduction designs using the split-unit principle is not quite uniform. The design we shall discuss now is known as a *split-block design* and also as a *split-plot design in strips*. It represents a variation of the simple split-plot design discussed in Section 13.2.

13.5.1 The Layout

The basic difference, and it is an important one, between the simple split-plot design and the split-block design is the way in which the levels of the two treatment factors are assigned to EUs. In this case both factors are applied to whole-plots which are

**Table 13.11 Outline of ANOVA for
SPD(RCBD, GRBD)**

Source	d.f.
Rep	$r - 1$
A	$a - 1$
Error(A)	$(r - 1)(a - 1)$
B	$b - 1$
$A \times B$	$(a - 1)(b - 1)$
Rep $\times B$	$(r - 1)(b - 1)$
Rep $\times A \times B$	$(r - 1)(a - 1)(b - 1)$
Error(B)	$rab(r' - 1)$
Total	$rar'b - 1$

“orthogonal” to each other. Schematically, this can be represented as follows, for factor A with levels $a_1, a_2, \dots, a_a (a = 8)$ and factor B with levels $b_1, b_2, \dots, b_b (b = 5)$:

		B				
		b_1	b_4	b_2	b_3	b_5
A	a_4	a_4b_1	a_4b_4	a_4b_2	a_4b_3	a_4b_5
	a_8	a_8b_1	a_8b_4	a_8b_2	a_8b_3	a_8b_5
	a_1	a_1b_1	a_1b_4	a_1b_2	a_1b_3	a_1b_5
	a_7	a_7b_1	a_7b_4	a_7b_2	a_7b_3	a_7b_5
	a_3	a_3b_1	a_3b_4	a_3b_2	a_3b_3	a_3b_5
	a_2	a_2b_1	a_2b_4	a_2b_2	a_2b_3	a_2b_5
	a_6	a_6b_1	a_6b_4	a_6b_2	a_6b_3	a_6b_5
	a_5	a_5b_1	a_5b_4	a_5b_2	a_5b_3	a_5b_5

An example of such an arrangement, where the levels of both factors are applied randomly to two types of whole-plots and the observations are obtained on the split-plots (determined by the intersection of the whole-plots) is the following agronomic experiment. We want to compare the yield of a certain crop under different systems of soil preparation and different density of seeding. Both operations (tilling and seeding) are done mechanically and it is impossible to perform both on small pieces of land. The arrangement shown above is then replicated r times, each time using different randomizations for A and B .

13.5.2 Linear Model and ANOVA

It is clear from our earlier discussion and from the nature of this arrangement, that we should have a separate error variance for comparisons among the levels of the A -factor, for comparisons among the levels of the B -factor, and for interaction comparisons. A model reflecting this structure is of the form

$$y_{ijk} = \mu + r_i + \alpha_j + e_{ij}^A + \beta_k + e_{ik}^B + (\alpha\beta)_{jk} + e_{ijk}^{AB} \quad (13.31)$$

with $i = 1, 2, \dots, r; j = 1, 2, \dots, a; k = 1, 2, \dots, b$ and the e_{ij}^A, e_{ik}^B , and e_{ijk}^{AB} can be considered as i.i.d. random variables with means 0 and variances $\sigma_{eA}^2, \sigma_{eB}^2$, and σ_{eAB}^2 , respectively. The ANOVA for this model is given in Table 13.12.

13.5.3 Estimating Treatment Contrasts

The ANOVA table suggests immediately how tests of hypotheses can be performed, using different error terms for tests about main effects and the interaction. Different error terms are also involved in obtaining the variances of estimable functions involving different kinds of treatment effects. We shall outline this briefly for the same comparisons discussed in Section 13.2:

(i) $\sum_j c_j \alpha_j$ is estimated by

$$\begin{aligned} \sum_j c_j \bar{y}_{.j} &= \frac{1}{rb} \sum_j c_j \left[\sum_{i,k} y_{ijk} \right] \\ &= \frac{1}{rb} \sum_j c_j \left[rb\mu + b \sum_i r_i + rb\alpha_j + b \sum_i e_{ij}^A + r \sum_k \beta_k + \sum_{i,k} e_{ik}^B \right. \\ &\quad \left. + r \sum_k (\alpha\beta)_{jk} + \sum_{i,k} e_{ijk}^{AB} \right] \\ &= \sum_j c_j \alpha_j + \frac{1}{r} \sum_{i,j} c_j e_{ij}^A + \frac{1}{rb} \sum_{i,j,k} c_j e_{ijk}^{AB} \end{aligned}$$

with

$$\begin{aligned} \text{var} \left(\sum_j c_j \hat{\alpha}_j \right) &= \frac{1}{r} \sum_j c_j^2 \sigma_{eA}^2 + \frac{1}{rb} \sum_j c_j^2 \sigma_{eAB}^2 \\ &= \frac{1}{rb} \sum_j c_j^2 [b\sigma_{eA}^2 + \sigma_{eAB}^2]. \end{aligned} \quad (13.32)$$

It follows then from Table 13.12 that

$$\hat{\text{var}} \left(\sum_j c_j \hat{\alpha}_j \right) = \frac{1}{rb} \sum_j c_j^2 \text{MS}(E_A) \quad (13.33)$$

Table 13.12 ANOVA for Split-Block Design

Source	d.f.	SS	MS	$E(MS)$
Replicates	$r - 1$	$ab \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$MS(R)$	
A factor	$a - 1$	$rb \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$	$MS(A)$	$\sigma_{\epsilon AB}^2 + b\sigma_{\epsilon A}^2 + rb \sum_j \alpha_j^2 / (a - 1)$
Error (A)	$(r - 1)(a - 1)$	$b \sum_{i,j} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	$MS(E_A)$	$\sigma_{\epsilon AB}^2 + b\sigma_{\epsilon A}^2$
B factor	$b - 1$	$ra \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$	$MS(B)$	$\sigma_{\epsilon AB}^2 + a\sigma_{\epsilon B}^2 + ra \sum_k \beta_k^2 / (b - 1)$
Error (B)	$(r - 1)(b - 1)$	$a \sum_{i,k} (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$	$MS(E_B)$	$\sigma_{\epsilon AB}^2 + a\sigma_{\epsilon B}^2$
A × B	$(a - 1)(b - 1)$	$r \sum_{j,k} (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$	$MS(AB)$	$\sigma_{\epsilon AB}^2 + r \sum_{j,k} (\alpha\beta)_{jk}^2 / (a - 1)(b - 1)$
Error (AB)	$(r - 1)(a - 1)(b - 1)$	$\sum_{i,j,k} (\bar{y}_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{i..} + \bar{y}_{..k} - \bar{y}_{...})^2$	$MS(EAB)$	$\sigma_{\epsilon AB}^2$
Total	$rab - 1$	$\sum_{i,j,k} (\bar{y}_{ijk} - \bar{y}_{...})^2$		

(ii) $\Sigma_k d_k \beta_k$ is estimated by $\Sigma_k d_k \bar{y}_{..k}$ with

$$\text{var} \left(\sum_k d_k \hat{\beta}_k \right) = \frac{1}{ra} \sum_k d_k^2 [a\sigma_{eB}^2 + \sigma_{eAB}^2] \quad (13.34)$$

following the arguments given for (13.32). It follows, again from Table 13.12, that

$$\text{var} \left(\sum_k d_k \hat{\beta}_k \right) = \frac{1}{ra} \sum_k d_k^2 \text{MS}(E_B) \quad (13.35)$$

(iii) $\Sigma_j c_j \alpha_j^{(p)}$ is estimated by

$$\begin{aligned} \sum_j c_j \left[\frac{1}{p} \sum_{k=1}^p \bar{y}_{.jk} \right] &= \frac{1}{rp} \sum_j c_j \left[\sum_i \sum_k' y_{ijk} \right] \left(\text{where } \sum_k' = \sum_{k=1}^p \right) \\ &= \frac{1}{rp} \sum_j c_j \left(rp\mu + p \sum_i r_i + rp\alpha_j + p \sum_i e_{ij}^A \right. \\ &\quad \left. + r \sum_k' \beta_k + \sum_i \sum_k' e_{ik}^B + r \sum_k' (\alpha\beta)_{jk} + \sum_i \sum_k' e_{ijk}^{AB} \right) \\ &= \sum_j c_j \alpha_j + \frac{1}{r} \sum_{ij} c_j e_{ij}^A + \frac{1}{p} \sum_j \sum_k' (\alpha\beta)_{jk} \\ &\quad + \frac{1}{rp} \sum_{i,j} \sum_k' c_j e_{ijk}^{AB} \end{aligned}$$

with

$$\begin{aligned} \text{var} \left(\sum_j c_j \hat{\alpha}_j^{(p)} \right) &= \frac{1}{r} \sum_j c_j^2 \sigma_{eA}^2 + \frac{1}{rp} \sum_j c_j^2 \sigma_{eAB}^2 \\ &= \frac{1}{rp} \sum_j c_j^2 [p\sigma_{eA}^2 + \sigma_{eAB}^2]. \end{aligned} \quad (13.36)$$

We know from Table 13.12 that

$$\hat{\sigma}_{eAB}^2 = \text{MS}(E_{AB})$$

and

$$\hat{\sigma}_{eA}^2 = \frac{1}{b} [\text{MS}(E_A) - \text{MS}(E_{AB})]$$

and hence we find the estimator for (13.36) to be

$$\hat{\text{var}} \left(\sum_j c_j \hat{\alpha}_j^{(p)} \right) = \frac{1}{rb} \sum_j c_j^2 \left[\text{MS}(E_A) + \frac{b-p}{p} \text{MS}(E_{AB}) \right]. \quad (13.37)$$

(iv) $\sum_k d_k \beta_k^{(q)}$ is estimated by $\sum_k d_k \left[\frac{1}{q} \sum_{j=1}^q \bar{y}_{.jk} \right]$ with

$$\text{var} \left(\sum_k d_k \hat{\beta}_k^{(q)} \right) = \frac{1}{rq} \sum_k d_k^2 [q\sigma_{eB}^2 + \sigma_{eAB}^2] \quad (13.38)$$

and

$$\hat{\text{var}} \left(\sum_k d_k \hat{\beta}_k^{(q)} \right) = \frac{1}{ra} \sum_k d_k^2 \left[\text{MS}(E_B) + \frac{a-q}{q} \text{MS}(E_{AB}) \right] \quad (13.39)$$

(v) $\tau_{jk} - \tau_{j'k'} (j \neq j', k \neq k')$ is estimated by $\bar{y}_{.jk} - \bar{y}_{.j'k'}$ with

$$\text{var}(\bar{y}_{.jk} - \bar{y}_{.j'k'}) = \frac{2}{r} (\sigma_{eA}^2 + \sigma_{eB}^2 + \sigma_{eAB}^2) \quad (13.40)$$

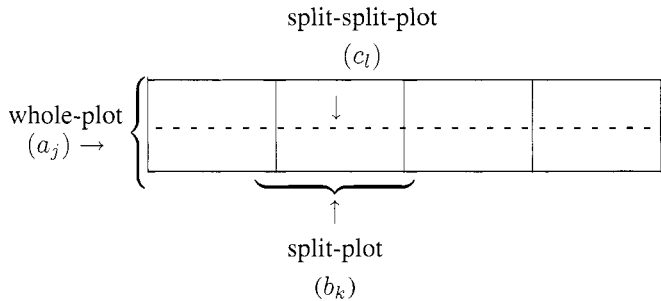
and

$$\hat{\text{var}}(\bar{y}_{.jk} - \bar{y}_{.j'k'}) = \frac{2}{r} \left[\frac{1}{b} \text{MS}(E_A) + \frac{1}{a} \text{MS}(E_B) + \left(1 - \frac{1}{a} - \frac{1}{b} \right) \text{MS}(E_{AB}) \right] \quad (13.41)$$

There can occur, obviously, also various forms of incomplete split-block designs. For example, we may have less than b column whole-plots. For a discussion of the analysis of such designs see Hering and Mejza (1997).

13.6 SPLIT-SPLIT-PLOT DESIGN

An extension of the simple split-plot design and its variations can be obtained by using the split-unit principle a second time, this time for the split-plots to obtain what are called *split-split-plots*:



This allows us to accommodate a third factor C with levels (split-split-plot treatments) c_1, c_2, \dots, c_c . Using three independent randomizations we assign the whole-plot treatments (a_j) to one whole-plot in each of r replicates, the split-plot treatments (b_k) to

Table 13.13 ANOVA for Split-Split-Plot Design

Source	d.f.	MS	$E(\text{MS})$
Replicates	$r - 1$	$\text{MS}(R)$	
A -factor	$a - 1$	$\text{MS}(A)$	$\sigma_{eC}^2 + c\sigma_{eB}^2 + bc\sigma_{eA}^2 + rbc \Sigma \alpha_j^2 / (a - 1)$
Error (A)	$(r - 1)(a - 1)$	$\text{MS}(E_A)$	$\sigma_{eC}^2 + c\sigma_{eB}^2 + bc\sigma_{eA}^2$
B -factor	$b - 1$	$\text{MS}(B)$	$\sigma_{eC}^2 + c\sigma_{eB}^2 + rac \Sigma \beta_k^2 / (b - 1)$
$A \times B$	$(a - 1)(b - 1)$	$\text{MS}(AB)$	$\sigma_{eC}^2 + c\sigma_{eB}^2 + rc \Sigma (\alpha\beta)_{jk}^2 / (a - 1)(b - 1)$
Error (B)	$(r - 1)a(b - 1)$	$\text{MS}(E_B)$	$\sigma_{eC}^2 + c\sigma_{eB}^2$
C -factor	$c - 1$	$\text{MS}(C)$	$\sigma_{eC}^2 + rab \Sigma \gamma_l^2 / (c - 1)$
$A \times C$	$(a - 1)(c - 1)$	$\text{MS}(AC)$	$\sigma_{eC}^2 + rb \Sigma (\alpha\gamma)_{jl}^2 / (a - 1)(c - 1)$
$B \times C$	$(b - 1)(c - 1)$	$\text{MS}(BC)$	$\sigma_{eC}^2 + ra \Sigma (\beta\gamma)_{kl}^2 / (b - 1)(c - 1)$
$A \times B \times C$	$(a - 1)(b - 1)(c - 1)$	$\text{MS}(ABC)$	$\sigma_{eC}^2 + r \Sigma (\alpha\beta\gamma)_{jkl}^2 / (a - 1)(b - 1)(c - 1)$
Error (C)	$(r - 1)ab(c - 1)$	$\text{MS}(E_C)$	σ_{eC}^2
Total	$rabc - 1$		

one split-plot within each whole-plot, and the split-split-plot treatments (c_l) to one split-split-plot within each split-plot. The observation for a split-split-plot is then the observation for the treatment combination $a_j b_k c_l$.

Extending the arguments used in Section 13.2 an appropriate model for observations from a split-split-plot experiment can be written as

$$\begin{aligned}
 y_{ijkl} = & \mu + r_i + \alpha_j + e_{ij}^A + \beta_k + (\alpha\beta)_{jk} + e_{ijk}^B + \gamma_l \\
 & + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} + e_{ijkl}^C.
 \end{aligned}
 \tag{13.42}$$

The error terms can be considered as i.i.d. random variables with means 0 and variances $\sigma_{eA}^2, \sigma_{eB}^2, \sigma_{eC}^2$, respectively. An outline of the ANOVA is given in Table 13.13.

The form of the $E(\text{MS})$ in Table 13.13 indicates how tests of hypotheses should be performed using the three different error terms. The number of different types of treatment comparisons becomes now quite large. The estimators and variances of such comparisons can be worked out easily using methods similar to those given in Sections 13.2 and 13.5. The estimated variances for some types of simple comparisons are given in Table 13.14. Here τ_{jkl} denotes the effect of the treatment combination $a_j b_k c_l$. Comparisons involving only the factors A and B are essentially as given in Section 13.2. For variances involving several MS the d.f. for a t -test have to be estimated using Satterthwaite's (1947) procedure (see also Section 13.2).

**Table 13.14 Estimated Variances of Treatment Comparisons
in a Split-Split-Plot Design**

Comparison	vâr
$\bar{\tau}_{..l} - \bar{\tau}_{..l'}$	$2MS(E_C)/rab$
$\bar{\tau}_{j..} - \bar{\tau}_{j'..}$	$2MS(E_C)/rb$
$\bar{\tau}_{.kl} - \bar{\tau}_{.kl'}$	$2MS(E_C)/ra$
$\tau_{jkl} - \tau_{jkl'}$	$2MS(E_C)/r$
$\bar{\tau}_{.kl} - \bar{\tau}_{.k'l'} (k \neq k')$	$2[(c-1)MS(E_C) + MS(E_B)]/rac$
$\tau_{jkl} - \tau_{jk'l}$	$2[(c-1)MS(E_C) + MS(E_B)]/rc$
$\bar{\tau}_{j..} - \bar{\tau}_{j'..} (j \neq j')$	$2[(c-1)MS(E_C) + MS(E_A)]/rbc$
$\tau_{jkl} - \tau_{j'kl}$	$2[b(c-1)MS(E_C) + (b-1)MS(E_B) + MS(E_A)]/rbc$

13.7 EXAMPLES USING SAS®

EXAMPLE 13.2: We consider here the simple split-plot design, the SPD(RCBD, RCBD), with $a = 3$ whole-plot treatments, $b = 2$ split-plot treatments and $r = 4$ replicates. The data are given in Table 13.15a.

To analyze the data we use both PROC GLM and PROC MIXED. The preferred procedure is PROC MIXED, but we include PROC GLM only for obtaining the ANOVA as given in Table 13.1. The input statements for both analyses are given in Table 13.15a:

- (i) The technical description for Error(A) is given by the (assumed to be negligible) interaction, rep*A.
- (ii) In the GLM analysis we have to specify the correct test for A by the test statement.
- (iii) In the MIXED analysis the rep*A interaction is declared to be random, thus enabling the correct test for A.
- (iv) In PROC MIXED we choose the Satterthwaite procedure to determine the correct d.f. for testing various hypotheses.

The results of both analyses are given in Table 13.15b:

- (v) In the type III ANOVA the P -values for rep, A, and rep*A should be ignored. The P -values for B and A * B are correct. The correct P -value for A (.0438) is given as a result of specifying the correct test (see (ii) above).

- (vi) Note that the d.f. for Error, our Error(B), are 9 and the d.f. for rep* A , our Error(A), are 6.
- (vii) In the MIXED analysis the tests for A , B , and $A * B$ are performed correctly, that is, with the correct error terms and the correct d.f. The results agree with those obtained with GLM.
- (viii) The d.f. for the three treatment comparisons specified in the input statement are given as 9, 9, 9.24, respectively. This agrees with our discussion in Section 13.2.5. □

Table 13.15 Split-Plot Design

a) Input statements:

```
data spltplot;
input rep A B y @@;
datalines;
1 1 1 56 1 1 2 41 1 2 1 50 1 2 2 36 1 3 1 39 1 3 2 35
2 1 1 36 2 1 2 25 2 2 1 36 2 2 2 28 2 3 1 33 2 3 2 30
3 1 1 32 3 1 2 24 3 2 1 31 3 2 2 27 3 3 1 15 3 3 2 19
4 1 1 30 4 1 2 25 4 2 1 35 4 2 2 30 4 3 1 17 4 3 2 18
run;

proc glm data=spltplot;
class rep A B;
model y = rep A rep*A B A*B;
test H=A E=rep*A;
title1 'SPD(RCBD, RCBD)';
title2 'BASIC ANOVA';
run;

proc mixed data=spltplot;
class rep A B;
model y = rep A B A*B/ddfm=Satterth;
random rep*A;
lsmeans A B A*B;
contrast 'a1+a2' vs a3' A 1 1 -2;
contrast 'b1-b2' B 1 -1;
estimate 'b1.b2' B 1 -1;
estimate 'a1b1-a1b2' B 1 -1 A*B 1 -1 0 0 0 0;
estimate 'a1b1-a2b1' A 1 -1 0 A*B 1 0 -1 0 0 0;
title2 'ANOVA RESULTS AND POST-HOC ANALYSIS';
run;
```

b) Output:

SPD(RCBD, RCBD)
BASIC ANOVA

The GLM Procedure

Class Level Information

Class	Levels	Values
-------	--------	--------

Table 13.15 (Continued)

	rep	4	1	2	3	4
	A	3	1	2	3	
	B	2	1	2		
	Number of Observations Read					24
	Number of Observations Used					24
Dependent Variable: y						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	14	2097.083333	149.791667	17.23	<.0001	
Error	9	78.250000	8.694444			
Corrected Total	23	2175.333333				
	R-Square	Coeff Var	Root MSE	y Mean		
	0.964029	9.460859	2.948634	31.16667		
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
rep	3	1241.000000	413.666667	47.58	<.0001	
A	2	353.083333	176.541667	20.31	0.0005	
rep*A	6	192.250000	32.041667	3.69	0.0394	
B	1	216.000000	216.000000	24.84	0.0008	
A*B	2	94.750000	47.375000	5.45	0.0281	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
rep	3	1241.000000	413.666667	47.58	<.0001	
A	2	353.083333	176.541667	20.31	0.0005	
rep*A	6	192.250000	32.041667	3.69	0.0394	
B	1	216.000000	216.000000	24.84	0.0008	
A*B	2	94.750000	47.375000	5.45	0.0281	
Tests of Hypotheses Using the Type III MS for rep*A as an Error Term						
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
A	2	353.083333	176.541667	5.51	0.0438	
SPD(RCBD, RCBD)						
ANOVA RESULTS AND POST-HOC ANALYSIS						
The Mixed Procedure						
Model Information						
Data Set	WORK.SPLTPLOT					
Dependent Variable	y					
Covariance Structure	Variance Components					
Estimation Method	REML					
Residual Variance Method	Profile					
Fixed Effects SE Method	Model-Based					
Degrees of Freedom Method	Satterthwaite					

Table 13.15 (Continued)

SPD(RCBD, RCBD)							
ANOVA RESULTS AND POST-HOC ANALYSIS							
The Mixed Procedure							
Covariance Parameter Estimates							
Cov Parm		Estimate					
rep*A		11.6736					
Residual		8.6944					
Fit Statistics							
-2 Res Log Likelihood		95.1					
AIC (smaller is better)		99.1					
AICC (smaller is better)		100.1					
BIC (smaller is better)		100.1					
Type 3 Tests of Fixed Effects							
Effect	Num DF	Den DF	F Value	Pr > F			
rep	3	6	12.91	0.0050			
A	2	6	5.51	0.0438			
B	1	9	24.84	0.0008			
A*B	2	9	5.45	0.0281			
Estimates							
Label	Estimate	Standard Error	DF	t Value	Pr > t		
b1_b2	6.0000	1.2038	9	4.98	0.0008		
alb1-alb2	9.7500	2.0850	9	4.68	0.0012		
alb1-a2b1	0.5000	3.1912	9.24	0.16	0.8789		
Contrasts							
Label	Num DF	Den DF	F Value	Pr > F			
(a1+a2) vs a3	1	6	10.99	0.161			
b1-b2	1	9	24.84	0.0008			
SPD(RCBD, RCBD)							
ANOVA RESULTS AND POST-HOC ANALYSIS							
The Mixed Procedure							
Least Square Means							
Effect	A	B	Estimate	Standard Error	DF	t value	Pr > t
A	1		33.6250	2.0013	6	16.80	<.0001

Table 13.15 (Continued)

A	2		34.1250	2.0013	6	17.05	<.0001
A	3		25.7500	2.0013	6	12.87	<.0001
B		1	34.1667	1.3028	9.24	26.23	<.0001
B		2	28.1667	1.3028	9.24	21.62	<.0001
A*B	1	1	38.5000	2.2565	9.24	17.06	<.0001
A*B	1	2	28.7500	2.2565	9.24	12.74	<.0001
A*B	2	1	38.0000	2.2565	9.24	16.84	<.0001
A*B	2	2	30.2500	2.2565	9.24	13.41	<.0001
A*B	3	1	26.0000	2.2565	9.24	11.52	<.0001
A*B	3	2	25.5000	2.2565	9.24	11.30	<.0001

EXAMPLE 13.3: Consider the SPD(CRD, RCBD) or between and within subjects design with $a = 3, b = 2$ with unequal numbers of subjects ($r_1 = 4, r_2 = 3, r_3 = 2$). The data are given in Table 13.16a together with the input statements for the analysis:

- (i) In the model statement we add the option “Satterth” in order to obtain the correct d.f. for treatment comparisons.
- (ii) Error(A) is specified as the random effect “subject(A)”.
- (iii) Error(A) and Error(B) both have 6 d.f.
- (iv) For the comparison “ $a_1b_1 - a_3b_1$ ” the d.f. are computed according to (13.23) as 7.99. □

Table 13.16 Between and Within Subjects Design

a) Input statements:

```
data spdcrd;
input A subject B y @@;
datalines;
1 1 1 25 1 1 2 28 1 2 1 27 1 2 2 31
1 3 1 28 1 3 2 32 1 4 1 30 1 4 2 32
2 5 1 30 2 5 2 35 2 6 1 33 2 6 2 39
2 7 1 33 2 7 2 35 3 8 1 49 3 8 2 55
3 9 1 49 3 9 2 54
;run;

proc mixed data=spdcrd;
class subject A B;
model y = A B A*B/ddfm=Satterth;
random subject(A);
lsmeans A B A*B;
estimate 'a1-a3' A 1 0 -1;
estimate 'b1-b2' B 1 -1;
estimate 'a1b1-a1b2' B 1 -1 A*B 1 -1 0 0 0 0;
```

Table 13.16 (Continued)

```
estimate 'alb1-a3b1' A 1 0 -1 A*B 1 0 0 0 -1 0;  
title1 'BETWEEN AND WITHIN SUBJECTS DESIGN';  
title2 'WITH UNEQUAL NUMBERS OF SUBJECTS';  
run;
```

b) Output:

BETWEEN AND WITHIN SUBJECTS DESIGN WITH UNEQUAL NUMBERS OF SUBJECTS										
The Mixed Procedure										
Model Information										
Data Set	WORK.SPDCRD									
Dependent Variable	Y									
Covariance Structure	Variance Components									
Estimation Method	REML									
Residual Variance Method	Profile									
Fixed Effects SE Method	Model-Based									
Degrees of Freedom Method	Satterthwaite									
Class Level Information										
Class	Levels	Values								
subject	9	1	2	3	4	5	6	7	8	9
A	3	1	2	3						
B	2	1	2							
Number of Observations										
Number of Observations Read										18
Number of Observations Used										18
Number of Observations Not Used										0
Covariance Parameter Estimates										
Cov Parm		Estimate								
subject (A)		2.4167								
Residual		0.9931								
Type 3 Tests of Fixed Effects										
Effect	Num DF	Den DF	F Value					Pr > F		
A	2	6	119.29					<.0001		
B	1	6	79.56					0.0001		
A*B	2	6	1.76					0.2511		
Estimates										
	Estimate	Standard Error	DF	t Value				Pr > t		
	-22.6250	1.4781	6	-15.31				<.0001		
	-4.3611	0.4889	6	-8.92				0.0001		
alb2	-3.2500	0.7046	6	-4.61				0.0036		
a3b1	-21.5000	1.5992	7.99	-13.44				<.0001		

Table 13.16 (Continued)

Least Squares Means							
Effect	A	B	Estimate	Standard Error	DF	t Value	Pr > t
A	1		29.1250	0.8534	6	34.13	<.0001
A	2		34.1667	0.9854	6	34.67	<.0001
A	3		51.7500	1.2069	6	42.88	<.0001
B		1	36.1667	0.6406	7.99	56.45	<.0001
B		2	40.5278	0.6406	7.99	63.26	<.0001
A*B	1	1	27.5000	0.9233	7.99	29.79	<.0001
A*B	1	2	30.7500	0.9233	7.99	33.31	<.0001
A*B	2	1	32.0000	1.0661	7.99	30.02	<.0001
A*B	2	2	36.3333	1.0661	7.99	34.08	<.0001
A*B	3	1	49.0000	1.3057	7.99	37.53	<.0001
A*B	3	2	54.5000	1.3057	7.99	41.74	<.0001

For the other split-plot designs mentioned in Section 13.4 we give below the input statements for PROC MIXED.

SPD(CRD, LSD) [Model (13.28)]:

```
CLASS      Subject  Order  A  B;
MODEL      Y = A    Order  B  A * B  A * Order/ddfm=satterth;
RANDOM      Subject(A);
```

SPD(LSD, RCBD) [Model (13.29)]:

```
CLASS      Row      Column  A  B;
MODEL      Y = Row   Column  A  B  A * B/ddfm=satterth;
RANDOM      Row*Column*A;
```

SPD(CRD, IBD): [Model (13.27)]:

```
CLASS      Subject  A  B;
MODEL      Y = A    B  A * B/ddfm=satterth;
RANDOM      Subject(A);
```

SPD(GRBD, RCBD) [Model (13.30)]:

```

CLASS      Block  A  Rep  B;
MODEL      Y = Block  A  Block * A  B  A * B
              Block * B  Block * A * B/ddfm=satterth;
RANDOM      Rep(A * Block);

```

where Rep refers to the replication of a treatment within a block.

SPD(RCBD, GRBD)

```

CLASS      Rep  A  B;
MODEL      Y = Rep  A  B  A * B  Rep*B/ddfm=satterth;
RANDOM      Rep*A;

```

SPLIT-BLOCK DESIGN [Model (13.31)]:

```

CLASS      Rep  A  B;
MODEL      Y = Rep  A  B  A * B /ddfm=satterth;
RANDOM      Rep*A  Rep*B;

```

SPLIT-SPLIT-PLOT DESIGN [Model (13.42)]:

```

CLASS      Rep  A  B;
MODEL      Y = Rep  A  B  A * B  C  A * C  B * C
              A * B * C/ddfm=satterth;
RANDOM      Rep*A  Rep*A * B;

```

13.8 EXERCISES

- 13.1** Consider the SPD(RCBD, RCBD) with a levels for the whole-plot factor, b levels for the split-plot factor, r replications, and subsampling, that is, n observations for each split-plot.

- (i) Write out an appropriate model for the observations.
- (ii) Write out the corresponding ANOVA table.
- (iii) Indicate how you would test hypotheses about the A -factor, the B -factor and the $A \times B$ interaction.
- (iv) Give the SAS commands for performing the analysis.
- (v) Using the notation of Section 13.2.4, obtain expressions for $\text{var}(\hat{\alpha}_j - \hat{\alpha}_{j'})$ and $\text{var}(\hat{\beta}_k - \hat{\beta}_{k'})$.

13.2 Consider the SPD (CRD, RCBD) and suppose that the A -factor itself has a factorial structure, that is, the a levels of A are the $a_1 \cdot a_2$ combinations of the a_1 levels of factor A_1 and the a_2 levels of factor A_2 . Similarly, the b levels of the B -factor are the $b_1 \cdot b_2$ combinations of the b_1 levels of factor B_1 and the b_2 levels of factor B_2 .

- (i) Write out a model for observations from this experiment.
- (ii) Write out the corresponding ANOVA table.
- (iii) Explain how you would test hypotheses about all main effects and interactions.
- (iv) Give the SAS commands for performing the analysis.

13.3 Suppose that in Exercise 13.2 the A - and B -factors are 2^2 factorials with factors A_1, A_2 and B_1, B_2 , respectively.

- (i) Give expressions for the estimates of the main effects A_1 and A_2 , and for the interaction $A_1 A_2$.
- (ii) Give expressions for $\text{var}(\hat{A}_1)$, $\text{var}(\hat{A}_2)$, $\text{var}(\widehat{A_1 A_2})$ and for the estimators of these variances.
- (iii) Do the same for $B_1, B_2, B_1 B_2$.
- (iv) Give an expression for the estimator for the interaction $A_1 B_1$, its variance, and the estimator for this variance.

13.4 Consider an experiment where the amount of dry matter is measured on wheat plants grown in different levels of moisture and with different fertilizers (Miliken and Johnson, 1984). The experimental material consists of 60 peat pots and 15 plastic trays; four (4) peat pots can be put in one tray. The moisture treatment consists of adding 10, 20, 30, 40, or 50 ml of water to the tray, where it will be absorbed by the pots. The experiment is being conducted at 3 different greenhouses such that 5 trays are used in each greenhouse and in each greenhouse each moisture level is assigned randomly to one tray. The fertilizer treatments are represented by a 2^2 factorial of low and high levels of nitrogen and phosphate. Each fertilizer combination is applied (at random) to individual pots in a tray such that each combination occurs once in each tray. In each pot 5 plants are grown, and observations are made on the individual plants.

- (i) Give a schematic (that is, graphical) representation of the layout of the experiment.
 - (ii) Give the name of the error-control design for this experiment.
 - (iii) Give an appropriate linear model for the design described in (ii), which reflects the structures of the error-control design, the treatment design, and the sampling design.
 - (iv) Outline the ANOVA table based on the model given in (iii), giving sources of variation, d.f., and $E(MS)$.
 - (v) Explain how you would test whether there exists interaction between nitrogen and the moisture treatment.
 - (vi) The researcher is interested in finding out whether there exists a linear trend for the effect of moisture on dry matter. Give an expression for the estimate of the linear trend and give its standard error.
- 13.5** Consider an experiment where the amount of dry matter is measured on wheat plants grown in different levels of moisture and with different fertilizers using a split-plot-type design (Milliken and Johnson, 1984). There are 48 different peat pots and 12 plastic trays; four (4) pots can be put in each tray. The moisture treatment consists of adding 40, 80, 120, or 160 ml of water to the tray, where it will be absorbed by the pots. The levels of moisture are assigned randomly to the trays such that each moisture level occurs 3 times. The fertilizer treatments are represented by all possible combinations of 0 and 1 unit of nitrogen, and 0 and 1 unit of phosphate. The fertilizer is applied individually to each pot in a tray such that each combination occurs once in each tray.
- (i) What are
 - (a) the whole-plots,
 - (b) the split-plots,
 - (c) the whole-plot treatment,
 - (d) the split-plot treatment?
 - (ii) What kind of split-plot-type design is this? Write out an appropriate linear model.
 - (iii) Outline the ANOVA table in as much detail as possible based on the description of the experiment (give sources of variation and d.f.).
 - (iv) Explain how you would test whether nitrogen has an effect on dry matter.
 - (v) The researcher is interested in finding out whether there exists a linear trend for the effect of moisture on dry matter. Give an expression for the estimate of the linear trend and its standard error (= square root of the estimated variance).
- 13.6** Discuss the layout and analysis of a SPD (BIBD, RCBD) and describe a possible application for this design.
- 13.7** Suppose a researcher comes to you to get some help on the analysis of the following data set:

Factor A	Factor B			
	b_1	b_2	b_3	b_4
a_1	x	x	x	x
	x	x	x	x
a_2	x	x	x	x
	x	x	x	x
a_3	x	x	x	x
	x	x	x	x

where each x represents an observation.

- (i) What questions would you ask the investigator before you can analyze the data?
- (ii) Describe three scenarios (analogous to Study 2 in Section 2.6.2) which could have given rise to this data set.
- (iii) For each scenario write out an appropriate linear model and the corresponding ANOVA table.
- (iv) For each case explain how you would make statistical inferences about the main effects A and B and the interaction $A \times B$.

CHAPTER 14

Designs with Repeated Measures

14.1 INTRODUCTION

For the title of this chapter we have, quite deliberately, not chosen the phrases *repeated measures designs* or *repeated measurement designs*, which, unfortunately, mean different things to different people. For Hedayat and Afsarinejad (1975), for example, they refer mainly to cross-over designs (see Section 10.7 and Chapter II.19), whereas for Hand and Crowder (1996), for example, they refer to designs with longitudinal data; that is, measurements repeated over time. This is the point of view we take here, too. In that sense then this aspect of experimental design is not so much an aspect of error-control or treatment design even though they play a role, as we shall see, but mainly an aspect of the observation design. As such repeated measures can be associated with any of the error-control designs we have discussed in previous chapters, for example a CRD with repeated measures.

We encounter repeated measures most often in medical, pharmaceutical, agricultural or psychological applications, where it is intended to study the efficacy of treatments over a certain time period.

EXAMPLE 14.1: (Frison and Pocock, 1992): A randomized trial of 152 patients with coronary heart disease compared an active drug with a placebo with respect to a possible adverse drug effect on the liver. The liver enzyme CPK was measured in each patient before treatment, at the time of randomization and every 1.5 months after treatment. □

14.2 METHODS FOR ANALYZING REPEATED MEASURES DATA

There exist several methods of analyzing such data. For an overview see Everitt (1995), and Keselman, Algina and Kowalchuk (2001). We shall mention here some methods and provide some more details in the following sections.

In order to keep the discussion simple let us consider the situation where t treatments are applied randomly to r experimental units (for instance, patients, animals, pieces of land), and measurements are being taken on each EU at p times after administration of the treatment, say t_1, t_2, \dots, t_p . In some situations a measurement at or immediately preceding the time of administration, say t_0 , may be taken. We, thus, have a CRD with p or $p + 1$ repeated measures. The time points may or may not be equidistant. The reader should have no difficulty extending the following discussion to other error-control designs. Also, Finney (1990) points out that repeated measures may not be confined to a temporal situation, but may involve also spatial situations as measurements are taken at different distances from the point of application of the treatment, for example, different depths of soil in a compaction study.

14.2.1 Comparisons at Separate Time Points

A commonly used approach is to consider the data at each time point arising from a separate “experiment”. Let us write a model for the observations as

$$y_{ijk} = \mu_{ijk} + e_{ijk} \quad (14.1)$$

with $i = 1, 2, \dots, t$; $j = 1, 2, \dots, r$; $k = 1, 2, \dots, p$ and

$$\mu_{ijk} = \mu + \tau_i + s_{ij} + T_k + (\tau T)_{ik} \quad (14.2)$$

with $\tau_i = i$ -th treatment effect

$s_{ij} =$ effect of the j -th subject (EU) for i -th treatment

$T_k = k$ -th time effect

$(\tau T)_{ik} =$ treatment-time interaction effect.

We should point out that model (14.2) is essentially equivalent to model (13.27) under the following correspondence:

$$\begin{aligned} \alpha_i &\rightarrow \tau_i \\ e_{ij}^A &\rightarrow s_{ij} \\ \beta_k &\rightarrow T_k \\ (\alpha\beta)_{ik} &\rightarrow (\tau T)_{ik} \\ e_{ijk}^B &\rightarrow e_{ijk} \end{aligned}$$

and the e_{ijk} have the following covariance structure. If we write

$$\mathbf{e}_{ij} = (e_{ij1}, e_{ij2}, \dots, e_{ijp})' \quad (14.3)$$

then

$$E(\mathbf{e}_{ij}) = \mathbf{0}$$

and

$$\text{var}(\mathbf{e}_{ij}) = \sum = (\sigma_{kk'}) \quad (14.4)$$

for all i, j , and $k, k' = 1, 2, \dots, p$.

For each time k we then consider contrasts of the form $\sum_i c_i \bar{\mu}_{i \cdot k}$ with $\sum c_i = 0$, by looking at the observations at each time point as arising from a CRD. We see from (14.2) that

$$\sum_i c_i \bar{\mu}_{i \cdot k} = \sum_i c_i (\tau_i + \bar{s}_{i \cdot} + (\tau T)_{ik}), \quad (14.5)$$

that is, the contrasts at different time points are possibly different because of treatment-time interactions. This is, of course, the main reason for looking at different time points as we want to find out whether the treatments have different effects over time, and if so, when the differences appear first.

A word of caution is in order here because the tests performed at each time point are not independent since the errors in (14.1) are now correlated. These correlations may become smaller as the time points are further apart. We may therefore choose time points which are not too "close" together, depending, of course, on the subject matter context.

14.2.2 Use of Summary Measures

Rather than performing several tests as described above, another approach may be to perform just one analysis based on a *summary measure* or *performance feature* for each subject over the entire set of time points. Such summary measures will have to be determined by the type of question we are investigating. For example, if we are interested in comparing the growth curve due to different treatments, then the area under the growth curve may be an appropriate summary measure. On other occasions the average response to treatment may be the most relevant summary measure (Matthews et al., 1990; Frison and Pocock, 1992).

14.2.3 Trend Analysis

In many situations it is important to detect trends over time or profiles or, perhaps even more importantly, to see whether the trends are the same for the different treatments. One way to approach these questions is as follows (see Rowell and Walters, 1976). Suppose we characterize the trends by a set of contrasts among the T_k in model (14.1), denoted by $\mathbf{c}'T$ with

$$\mathbf{c}_l = (c_{1l}, c_{2l}, \dots, c_{pl})' \quad (14.6)$$

Table 14.1 ANOVA for Model (14.7)

Source	d.f.	SS	$E(MS)$
μ_l^*	1	$tr \bar{z}_{l..}^2$	$\sigma_l^2 + tr \mu_l^{*2}$
γ_{li}^*	$t - 1$	$r \sum_{i=1}^t (\bar{z}_{li.} - \bar{z}_{l..})^2$	$\sigma_l^2 + r \frac{\sum_i \gamma_{li}^{*2}}{t - 1}$
Error _{l}	$t(r - 1)$	$\sum_{i=1}^t \sum_{j=1}^r (z_{lij} - \bar{z}_{li.})^2$	σ_l^2
Total _{l}	tr	$\sum_{i=1}^t \sum_{j=1}^r z_{lij}^2$	

and

$$\mathbf{c}_l' \mathcal{J} = 0 \quad (l = 1, 2, \dots, q),$$

and

$$\mathbf{T} = (T_1, T_2, \dots, T_p)'$$

In many cases the time points will be equally spaced, for instance, in intervals of 15 minutes, in which case it is useful and convenient to characterize the trend as a polynomial over time and take the \mathbf{c}_l of (14.6) as the orthogonal polynomials of order p and degree $l = 1, 2, \dots, q$ (see Chapter 7). From (14.1) we then derive sequentially the models (that is, for $l = 1, 2, \dots, q$)

$$\begin{aligned}
 z_{lij} &= \sum_k c_{kl} y_{ijk} = \sum_k c_{kl} T_k + \sum_k c_{kl} (\tau T)_{ik} + \sum_k c_{kl} e_{ijk} \\
 &= \mu_i^* + \gamma_{li}^* + e_{lij}^*
 \end{aligned} \tag{14.7}$$

with

$$\begin{aligned}
 \mu_i^* &= \sum_k c_{kl} T_k \\
 \gamma_{li}^* &= \sum_k c_{kl} (\tau T)_{ik} \\
 e_{lij}^* &= \sum_k c_{kl} e_{ijk}
 \end{aligned}$$

and $E(e_{lij}^*) = 0$, $\text{var}(e_{lij}^*) = \mathbf{c}_l' \Sigma \mathbf{c}_l = \sigma_l^2$ say. Model (14.7) leads to the ANOVA given in Table 14.1.

Hence, to test for an overall trend defined by (14.4), that is, $H_0: \sum_k c_{kl}T_k = \mu_l^* = 0$, we use the test

$$F = \frac{SS(\mu_l^*)}{MS(\text{Error}_l)}$$

with 1 and $a(r-1)$ d.f., and to test whether this trend is the same for all the treatments, that is, $H_0: \gamma_{l1}^* = \gamma_{l2}^* = \dots = \gamma_{lt}^*$, we use the test statistic

$$F = \frac{MS(\gamma_{li}^*)}{MS(\text{Error}_l)}$$

with $a-1$ and $a(r-1)$ d.f. We perform this analysis for every $l = 1, 2, \dots, q$. These tests generally provide an informative picture about the behavior of the treatments over time.

14.2.4 The ANOVA Method

In Section 13.4.2 we have referred to the design with repeated measures as a split-plot design in time. We have pointed out, however, that there is no randomization for the B -factor, that is, for time, and that the $e_{ijk}^B = e_{ijk}$ in model (13.27) and (14.1), respectively, have a correlation structure, which we now have acknowledged explicitly in (14.4). For these reasons the testing procedures derived from the ANOVA in Table 13.3 may be invalid.

However, if \sum in (14.4) satisfies the so-called Huynh-Feldt condition (Huynh and Feldt, 1970) given as

$$\sum = \lambda \mathbf{I}_p + \gamma \mathbf{J}'_p + \mathbf{J}_p \gamma', \quad (14.8)$$

where λ is a constant and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)'$ is a vector of constants, then the usual F -test for testing $H_0: T_1 = T_2 = \dots = T_p$ [model (13.27) and 14.1] is valid. The condition (14.8) which can be written alternatively as

$$\sigma_{kk'} = \lambda \delta_{kk'} + \gamma_k + \gamma_{k'},$$

where $\delta_{kk'} = 1$ if $k = k'$ and $= 0$ otherwise, contains as a special case a structure referred to as *compound symmetry*, characterized by

$$\sigma_{kk'} = \begin{cases} \sigma^2 & \text{for } k = k' \\ \rho\sigma^2 & \text{for } k \neq k' \end{cases}, \quad (14.9)$$

that is, equal variances and covariances for the $e_{ijk}^{(B)}$. For the case of compound symmetry (CS) for the $e_{ij}^{(B)}$ in (14.3) Geisser and Greenhouse (1958) already proved that the usual analysis for the split-plot design is valid. The case we need to consider here then in connection with repeated measures designs is when neither (14.9) nor (14.8) are satisfied.

14.2.5 Mixed Model Analysis

Recall that the data from a CRD with repeated measures (often also referred to as a between- and within-subjects design) are described by the model (see (14.1), (14.2), (14.4))

$$y_{ijk} = \mu + \tau_k + s_{ij} + T_k + (\tau T)_{ik} + e_{ijk}, \quad (14.10)$$

where $\mu + \tau_i + T_k + (\tau T)_{ik}$ is the fixed part and s_{ij} is the random part of a mixed model. More specifically, concerning the random part, the s_{ij} are i.i.d. $(0, \sigma_s^2)$ random variables, and the e_{ijk} have a covariance structure given by (14.4). As a consequence the variance-covariance matrix for the vector of $n = trp$ observations, \mathbf{y} , is given by

$$\text{var}(\mathbf{y}) = \mathbf{I}_n \sigma_s^2 + \mathbf{I}_{tr} \times \sum \equiv \mathbf{V}, \quad (14.11)$$

where “ \times ” indicates a Kronecker product.

It is \mathbf{V} of (14.11) that we would need to use to estimate and make inference about the fixed effects in model (14.10) (see Sections 4.6.2 and 4.18). Unfortunately, we do not know the variance and covariance components in (14.11) and, indeed, we do not even know the covariance structure represented by \sum . Hence, in order to analyze repeated measures data we need to make an assumption about the structure \sum and then use a suitable estimation procedure to estimate the variance and covariance component to obtain $\widehat{\sum}$, say, and then solve the Aitken-like equations (4.80) using $\widehat{\sum}$. This is, generally speaking, not an easy task and for the average user possible only with the availability of suitable software, such as SAS PROC MIXED (SAS Institute, 2002-2003).

We shall not go into the details of SAS PROC MIXED but mention the form of some of the possible covariance structures (using $p = 4$) that this program can use and that we consider to be relevant for this situation:

Compound Symmetry (CS):

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

that is, all the variances (diagonal) are the same, and all covariances (off-diagonal) are the same, regardless of the distance between time points (see (14.9)).

First-order Autoregressive (AR(1)):

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

that is, the variances are the same, and, since $|\rho| < 1$, the covariance diminishes as the time points become further apart.

Unstructured (UN):

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

that is, all variances and all covariances are possibly different.

Spatial Power (SP(POW)(C)):

$$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix}$$

that is, just like AR(1) the correlations depend on the distance between time or spatial points, except that the power of ρ is now determined by a measure, d_{ij} , of the actual distance between points i and j .

As we have mentioned before, it is unlikely that the CS is appropriate for repeated measures data, but if this structure holds then the analysis is equivalent to the ANOVA given in Table 13.3. This is the reason why the CS structure is appealing and frequently used. Even though we do not go as far as Finney (1990) who says that it should never be used (unless $p = 2$), we caution the user to be very careful with its use.

We prefer, in general, the AR(1) structure because it seems to reflect an intuitive amount of correlation between observations at different time points and to allow for the correlation to become smaller as the times of observation are farther apart. The same comments apply to SP(POW)(C), in particular if the distances between points are not the same.

The safest assumption about Σ is certainly UN. But the drawback is that it requires the estimation of many parameters in \mathbf{V} which will make it not a very powerful procedure.

SAS PROC MIXED allows for other covariance structures, but the ones mentioned above will generally suffice from a practical point of view, and we shall illustrate their use in Section 14.3.

To conform somewhat to the SAS PROC MIXED notation we shall rewrite model (14.1) for the general situation (in matrix notation) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}\boldsymbol{\beta} + \mathbf{e}, \quad (14.12)$$

where $\mathbf{X}\alpha$ refers to the fixed part, and $\mathbf{U}\beta$ refers to the random part, with \mathbf{X} and \mathbf{U} known matrices and

$$\begin{aligned} E(\beta) &= \mathbf{0}, & \text{var}(\beta) &= \mathbf{G} \\ E(e) &= \mathbf{0}, & \text{var}(e) &= \mathbf{R} \end{aligned} \quad (14.13)$$

so that (14.11) becomes

$$\text{var}(\mathbf{y}) = \mathbf{UGU}' + \mathbf{R} \equiv \mathbf{V}. \quad (14.14)$$

In our situation the fixed part of (14.12) represents parameters associated with the treatment, time, and treatment \times time interaction effects as well as possibly some blocking factor effects for error-control designs other than the CRD. In addition, we may also have other treatment effects in connection with a real split-plot structure, which would then contribute further error terms to $\mathbf{U}\beta$. Thus model (14.12) represents the most general case.

The fitting of model (14.12) in SAS PROC MIXED can be done by specifying one of several procedures. The default option is the residual maximum likelihood procedure (REML). For a description see Section II. 1.11.2.

14.3 EXAMPLES USING SAS®

EXAMPLE 14.2: Consider an experiment comparing different drugs with respect to their efficacy to control the heart rate of certain patients. We have $t = 3$ drugs, each drug being given to $r = 5$ patients, and the heart rate is measured at $p = 4$ different (equispaced) times. The data are given in Table 14.2a.

We perform several analysis mainly to illustrate different procedures as described in Section 14.2.4 and 14.2.5 and show their similarities and differences.

The SAS PROC GLM and MIXED for the ANOVA method and the mixed model analysis using the covariance structures CS, AR(1), and UN are given in Table 14.2a. In each case we perform the basic analysis. Only for the AR(1) method (which is our preferred method) do we follow up with a post-hoc analysis. The results for all analysis are given in Table 14.2b.

We comment as follows:

- (i) For the ANOVA method we need to specify the correct error term for testing hypotheses about drugs. This error term corresponds to error(A) in the SPD(CRD, RCBD) (see Section 13.4.1) and is given in technical terms by “person(drug)”.
- (ii) The results show that there are no significant differences among drugs ($P = .29$), but we shall return to this point later in light of the fact that there are significant changes over time ($P < .0001$) and, most importantly, significant interaction between drugs and time ($P < .0001$).
- (iii) For the CS method we obtain the same test results as mentioned in (ii) above.

Table 14.2 CRD with Repeated Measures

a) Input statements:**data heart;**

input drug person time rate @@;

datalines;

```

1 1 1 72 1 1 2 86 1 1 3 81 1 1 4 77
1 2 1 78 1 2 2 83 1 2 3 88 1 2 4 81
1 3 1 71 1 3 2 82 1 3 3 81 1 3 4 75
1 4 1 72 1 4 2 83 1 4 3 83 1 4 4 69
1 5 1 66 1 5 2 79 1 5 3 77 1 5 4 66
2 1 1 85 2 1 2 86 2 1 3 83 2 1 4 80
2 2 1 82 2 2 2 86 2 2 3 80 2 2 4 84
2 3 1 71 2 3 2 78 2 3 3 70 2 3 4 75
2 4 1 83 2 4 2 88 2 4 3 79 2 4 4 81
2 5 1 86 2 5 2 85 2 5 3 76 2 5 4 76
3 1 1 69 3 1 2 73 3 1 3 72 3 1 4 74
3 2 1 66 3 2 2 62 3 2 3 67 3 2 4 73
3 3 1 84 3 3 2 90 3 3 3 88 3 3 4 87
3 4 1 80 3 4 2 81 3 4 3 77 3 4 4 72
3 5 1 72 3 5 2 72 3 5 3 69 3 5 4 70

```

```

;
run;

```

proc glm data=heart;

class drug person time;

model rate=drug person(drug) time drug*time/SS3;

test h=drug e=person(drug);

title1 'CRD WITH REPEATED MEASURES';

title2 'ANALYZED AS SPD(CRD,RCBD)';

run;**proc mixed data=heart;**

class drug person time;

model rate=drug time drug*time;

repeated/type=cs subject=person(drug) rcorr;

title2 'WITH COMPOUND SYMMETRY';

run;**proc mixed data=heart;**

class drug person time;

model rate=drug time drug*time;

repeated/type=ar(1) subject=person(drug) rcorr;

estimate 'drug1lin' time -3 -1 1 3 drug*time -3 -1 1 3 0 0 0 0 0 0 0;

estimate 'drug2lin' time -3 -1 1 3 drug*time 0 0 0 -3 -1 1 3 0 0 0 0;

estimate 'drug3lin' time -3 -1 1 3 drug*time 0 0 0 0 0 0 0 -3 -1 1 3;

estimate 'drug1qua' time 1 -1 -1 1 drug*time 1 -1 -1 1 0 0 0 0 0 0 0;

estimate 'drug2qua' time 1 -1 -1 1 drug*time 0 0 0 1 -1 -1 1 0 0 0 0;

estimate 'drug3qua' time 1 -1 -1 1 drug*time 0 0 0 0 0 0 0 1 -1 -1 1;

estimate 'drug1cub' time -1 3 -3 1 drug*time -1 3 -3 1 0 0 0 0 0 0 0;

estimate 'drug2cub' time -1 3 -3 1 drug*time 0 0 0 -1 3 -3 1 0 0 0 0;

estimate 'drug3cub' time -1 3 -3 1 drug*time 0 0 0 0 0 0 0 -1 3 -3 1;

title2 'WITH AUTOREGRESSIVE ERRORS';

run;

Table 14.2 (Continued)

```
proc mixed data=heart;
class drug time;
model rate=drug time drug*time;
repeated/type=un subject=person(drug) rcorr;
title2 'WITH UNSPECIFIED CORRELATION STRUCTURE';
run;

proc sort data=heart;
by drug time;
proc means mean noprint;
by drug time;
var rate;
output out=meandata mean=m_rate;
title2 'PLOT OF MEAN HEART RATE OVER TIME';
run;
proc plot data=meandata;
plot m_rate*time=drug;
label m_rate='Mean Heart Rate';
run;

quit;
```

b.) Output:

CRD WITH REPEATED MEASURES		
ANALYZED AS SPD (CRD,RCBD)		
The GLM Procedure		
Class Level Information		
Class	Levels	Values
drug	3	1 2 3
person	5	1 2 3 4 5
time	4	1 2 3 4
Number of Observations Read		60
Number of Observations Used		60

Dependent Variable: rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	2449.500000	106.500000	12.73	<.0001
Error	36	301.100000	8.363889		
Corrected Total	59	2750.600000			

Table 14.2 (Continued)

R-Square	Coeff Var	Root MSE	rate	Mean
0.890533	3.722058	2.892039	77.70000	

Source	DF	Type III SS	Mean Square	F Value	Pr > F
drug	2	337.600000	168.800000	20.18	<.0001
person(drug)	12	1498.500000	124.875000	14.93	<.0001
time	3	256.333333	85.444444	10.22	<.0001
drug*time	6	357.066667	59.511111	7.12	<.0001

Tests of Hypotheses Using the Type III
MS for person(drug) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
drug	2	337.600000	168.800000	1.35	0.2955

The Mixed Procedure

Model Information

Data Set	WORK.HEART
Dependent Variable	rate
Covariance Structure	Compound Symmetry
Subject Effect	person(drug)
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	329.48905107	
1	1	289.92035887	0.00000000

Convergence criteria met.

CRD WITH REPEATED MEASURES
WITH COMPOUND SYMMETRY

The Mixed Procedure

Estimated R Correlation Matrix for person(drug) 1 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.7769	0.7769	0.7769
2	0.7769	1.0000	0.7769	0.7769
3	0.7769	0.7769	1.0000	0.7769
4	0.7769	0.7769	0.7769	1.0000

Table 14.2 (Continued)

Covariance Parameter Estimates				
Cov Parm	Subject		Estimate	
CS	person(drug)		29.1278	
Residual			8.3639	
Fit Statistics				
-2 Res Log Likelihood			289.9	
AIC (smaller is better)			293.9	
AICC (smaller is better)			294.2	
BIC (smaller is better)			295.3	
Null Model Likelihood Ratio Test				
DF	Chi-Square	Pr > ChiSq		
1	39.57	<.0001		
Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	12	1.35	0.2955
time	3	36	10.22	<.0001
drug*time	6	36	7.12	<.0001
CRD WITH REPEATED MEASURES WITH AUTOREGRESSIVE ERRORS				
The Mixed Procedure				
Model Information				
Data Set	WORK.HEART			
Dependent Variable	rate			
Covariance Structure	Autoregressive			
Subject Effect	person(drug)			
Estimation Method	REML			
Residual Variance Method	Profile			
Fixed Effects SE Method	Model-Based			
Degrees of Freedom Method	Between-Within			
Iteration History				
Iteration	Evaluations	-2 Res Log Like	Criterion	
0	1	329.48905107		
1	2	285.94895046	0.00006372	
2	1	285.94254325	0.00000004	
3	1	285.94253892	0.00000000	

Convergence criteria met.

Table 14.2 (Continued)

Estimated R Correlation Matrix for person(drug) 1 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.8278	0.6852	0.5672
2	0.8278	1.0000	0.8278	0.6852
3	0.6852	0.8278	1.0000	0.8278
4	0.5672	0.6852	0.8278	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
AR(1)	person(drug)	0.8278
Residual		36.0107

Fit Statistics

-2 Res Log Likelihood	285.9
AIC (smaller is better)	289.9
AICC (smaller is better)	290.2
BIC (smaller is better)	291.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	43.55	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
drug	2	12	1.46	0.2707
time	3	36	14.53	<.0001
drug*time	6	36	8.57	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
drug1lin	4.8000	8.4206	36	0.57	0.5722
drug2lin	-13.6000	8.4206	36	-1.62	0.1150
drug3lin	2.0000	8.4206	36	0.24	0.8136
drug1qua	-19.2000	2.3054	36	-8.33	<.0001
drug2qua	-1.6000	2.3054	36	-0.69	0.4921
drug3qua	-0.8000	2.3054	36	-0.35	0.7306
drug1cub	3.6000	4.0298	36	0.89	0.3776
drug2cub	18.8000	4.0298	36	4.67	<.0001
drug3cub	4.0000	4.0298	36	0.99	0.3275

CRD WITH REPEATED MEASURES
WITH UNSPECIFIED CORRELATION STRUCTURE

The Mixed Procedure

Table 14.2 (Continued)

Model Information				
Data Set	WORK.HEART			
Dependent Variable	rate			
Covariance Structure	Unstructured			
Subject Effect	person(drug)			
Estimation Method	REML			
Residual Variance Method	None			
Fixed Effects SE Method	Model-Based			
Degrees of Freedom Method	Between-Within			

Iteration History				
Iteration	Evaluations	-2 Res Log Like	Criterion	
0	1	329.48905107		
1	1	278.84809316	0.00000000	

Estimated R Correlation Matrix for person(drug) 1 1				
Row	Col1	Col2	Col3	Col4
1	1.0000	0.8498	0.8889	0.6246
2	0.8498	1.0000	0.8697	0.6315
3	0.8889	0.8697	1.0000	0.7945
4	0.6246	0.6315	0.7945	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	person(drug)	37.2333
UN(2,1)	person(drug)	34.3167
UN(2,2)	person(drug)	43.8000
UN(3,1)	person(drug)	32.9333
UN(3,2)	person(drug)	34.9500
UN(3,3)	person(drug)	36.8667
UN(4,1)	person(drug)	21.5833
UN(4,2)	person(drug)	23.6667
UN(4,3)	person(drug)	27.3167
UN(4,4)	person(drug)	32.0667

Fit Statistics	
-2 Res Log Likelihood	278.8
AIC (smaller is better)	298.8
AICC (smaller is better)	304.8
BIC (smaller is better)	305.9

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
9	50.64	<.0001

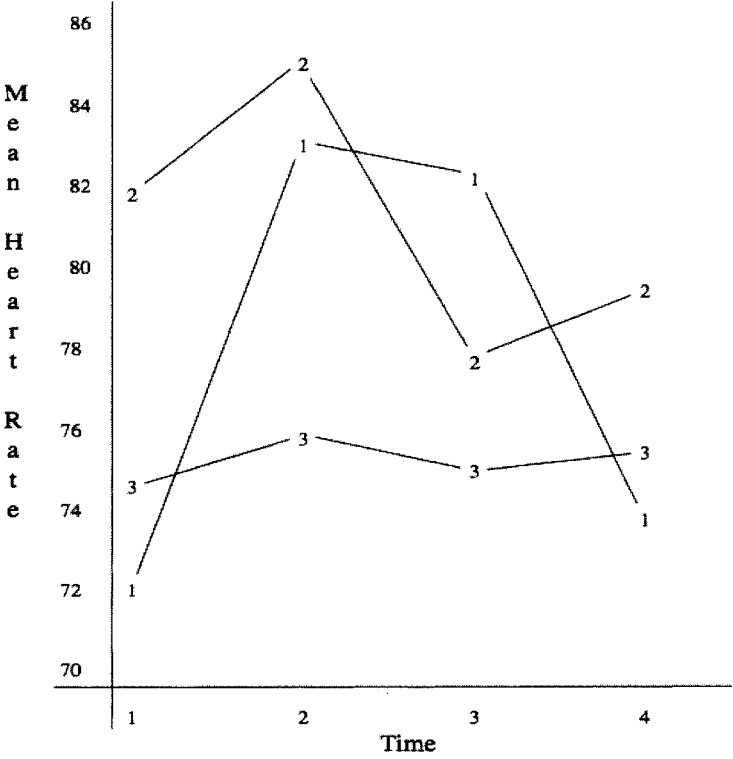
Table 14.2 (Continued)

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
drug	2	12	1.35	0.2955
time	3	12	12.35	0.0006
drug*time	6	12	17.31	<.0001

CRD WITH REPEATED MEASURES
PLOT OF MEAN HEART RATE OVER TIME

Plot of m_rate*time. Symbol is value of drug.



- (iv) The option “rcorr” in the input statements for all mixed model procedures results in printing out the correlation matrix for one person (subject).
- (v) For the CS method the correlation matrix is given by $.2231\mathbf{J} + .7769\mathbf{J}\mathbf{J}'$, that is, the correlation between two observations for the same person is estimated as $r = .7769$. This value is obtained as the intra-class correlation

$$r = \frac{\hat{\sigma}_{e(A)}^2}{\hat{\sigma}_{e(A)}^2 + \hat{\sigma}_{e(B)}^2},$$

where

$$\begin{aligned}\hat{\sigma}_{e(B)}^2 &= (\text{MS}[\text{Person}(\text{drug})] - \text{MS}(E))/4 \\ &= (124.875 - 8.364)/4 = 29.1278.\end{aligned}$$

- (vii) The correlation between the observations at two adjacent time points is estimated as $r = .8278$.
- (viii) For the UN method the test results, again, are essentially the same as for the other analyses.
- (ix) The estimated correlation structure, in fact, shows that at least for the first three time points the correlations are almost equal, that is, exhibiting a CS structure.
- (x) Looking at the drug-time interaction plot it seems worthwhile to look at the individual trends. To do that we fit linear, quadratic and cubic polynomials. The results show that there is a significant quadratic trend for drug 1 ($P < .0001$) and a significant cubic trend for drug 2 ($P < .0001$). The interaction is clearly not codirectional and hence it may not be appropriate to consider a test for the overall drug effects (see (ii) above). \square

EXAMPLE 14.3: The basic design for this pollution study is a SPD(CRD, RCBD) (see Section 13.4.1). We have $a = 2$ pollutants (P) as whole-plot treatments, $r = 2$ replications, $b = 4$ split-plot treatments with a 2^2 factorial structure (two varieties, V_1 , V_2 , and two growth enhancing treatments, A_1 , A_2 ,). The pollutants are applied to four pots in a growth chamber (CH), each pot containing a plant from either V_1 or V_2 treated with either A_1 or A_2 such that all four combinations of (V_i, A_j) are represented in each growth chamber. Each pollutant is assigned randomly to two growth chambers. The four pots are randomly arranged in the growth chambers. Each plant is measured at three times, the measurements constituting the repeated measures. The data are given in Table 14.3a:

- (i) We include in the model statement the main effect P, V, A, TIME and all interactions among the corresponding factors up to three-factor interactions (as indicated by “@3”).

Table 14.3 Repeated Measures in SPD(CRD, RCBD)**a) Input statements:**

```

data pollutn;
input P CH V A TIME PLANT y @@;
datalines;
1 1 1 1 1 23 1 1 1 1 2 1 24 1 1 1 1 3 1 26
1 1 1 2 1 20 1 1 1 2 2 2 24 1 1 1 2 3 2 25
1 1 2 1 1 3 26 1 1 2 1 2 3 29 1 1 2 1 3 3 33
1 1 2 2 1 4 31 1 1 2 2 2 4 35 1 1 2 2 3 4 38
1 2 1 1 1 5 25 1 2 1 1 2 5 26 1 2 1 1 3 5 27
1 2 1 2 1 6 30 1 2 1 2 2 6 35 1 2 1 2 3 6 36
1 2 2 1 1 7 28 1 2 2 1 2 7 30 1 2 2 1 3 7 34
1 2 2 2 1 8 32 1 2 2 2 2 8 33 1 2 2 2 3 8 36
2 3 1 1 1 9 40 2 3 1 1 2 9 43 2 3 1 1 3 9 45
2 3 1 2 1 10 44 2 3 1 2 2 10 47 2 3 1 2 3 10 48
2 3 2 1 1 11 48 2 3 2 1 2 11 50 2 3 2 1 3 11 55
2 3 2 2 1 12 52 2 3 2 2 2 12 57 2 3 2 2 3 12 60
2 4 1 1 1 13 45 2 4 1 1 2 13 47 2 4 1 1 3 13 50
2 4 1 2 1 14 45 2 4 1 2 2 14 49 2 4 1 2 3 14 52
2 4 2 1 1 15 56 2 4 2 1 2 15 57 2 4 2 1 3 15 60
2 4 2 2 1 16 53 2 4 2 2 2 16 57 2 4 2 2 3 16 59
;
run;

proc print data=pollutn;
title1 'POLLUTION DATA';
run;

proc mixed data=pollutn;
class P CH V A TIME PLANT;
model y=P|V|A|TIME @3/ddfm=satterth;
random CH(P) V*A*CH(P);
repeated/type=ar(1) subject=PLANT(P*V*A) rcorr;
lsmeans P V A TIME V*TIME A*TIME;
title2 'REPEATED MEASURES ANALYSIS';
title3 'WITH AUTOCORRELATED ERRORS';
run;

```

Table 14.3 (Continued)

b.) Output:

POLLUTION DATA							
Obs	P	CH	V	A	TIME	PLANT	y
1	1	1	1	1	1	1	23
2	1	1	1	1	2	1	24
3	1	1	1	1	3	1	26
4	1	1	1	2	1	2	20
5	1	1	1	2	2	2	24
6	1	1	1	2	3	2	25
7	1	1	2	1	1	3	26
8	1	1	2	1	2	3	29
9	1	1	2	1	3	3	33
10	1	1	2	2	1	4	31
11	1	1	2	2	2	4	35
12	1	1	2	2	3	4	38
13	1	2	1	1	1	5	25
14	1	2	1	1	2	5	26
15	1	2	1	1	3	5	27
16	1	2	1	2	1	6	30
17	1	2	1	2	2	6	35
18	1	2	1	2	3	6	36
19	1	2	2	1	1	7	28
20	1	2	2	1	2	7	30
21	1	2	2	1	3	7	34
22	1	2	2	2	1	8	32
23	1	2	2	2	2	8	33
24	1	2	2	2	3	8	36
25	2	3	1	1	1	9	40
26	2	3	1	1	2	9	43
27	2	3	1	1	3	9	45
28	2	3	1	2	1	10	44
29	2	3	1	2	2	10	47
30	2	3	1	2	3	10	48
31	2	3	2	1	1	11	48
32	2	3	2	1	2	11	50
33	2	3	2	1	3	11	55
34	2	3	2	2	1	12	52
35	2	3	2	2	2	12	57
36	2	3	2	2	3	12	60
37	2	4	1	1	1	13	45
38	2	4	1	1	2	13	47
39	2	4	1	1	3	13	50
40	2	4	1	2	1	14	45
41	2	4	1	2	2	14	49
42	2	4	1	2	3	14	52
43	2	4	2	1	1	15	56
44	2	4	2	1	2	15	57
45	2	4	2	1	3	15	60
46	2	4	2	2	1	16	53
47	2	4	2	2	2	16	57
48	2	4	2	2	3	16	59

Table 14.3 (Continued)

POLLUTION DATA															
REPEATED MEASURES ANALYSIS															
WITH AUTOCORRELATED ERRORS															
The Mixed Procedure															
Model Information															
Data Set	WORK.POLLUTN														
Dependent Variable	Y														
Covariance Structures	Variance Components, Autoregressive														
Subject Effect	PLANT(P*V*A)														
Estimation Method	REML														
Residual Variance Method	Profile														
Fixed Effects SE Method	Model-Based														
Degrees of Freedom Method	Satterthwaite														
Class Level Information															
Class	Levels	Values													
P	2	1 2													
CH	4	1 2 3 4													
V	2	1 2													
A	2	1 2													
TIME	3	1 2 3													
PLANT	16	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16													
Iteration History															
Iteration	Evaluations	-2 Res Log Like												Criterion	
0	1	157.97668907													
1	3	115.02581938												0.02422314	
2	3	114.57079107												0.00015995	
3	1	114.56527273												0.00000045	
4	1	114.56525763												0.00000000	
Convergence criteria met.															
Estimated R Correlation Matrix for PLANT(P*V*A) 1 1 1 1															
Row	Col1	Col2	Col3												
1	1.0000	0.9355	0.8751												
2	0.9355	1.0000	0.9355												
3	0.8751	0.9355	1.0000												
Covariance Parameter Estimates															
Cov Parm	Subject											Estimate			
CH(P)												3.5989			
CH*V*A(P)												1.66E-15			
AR(1)	PLANT(P*V*A)											0.9355			
Residual												8.8660			

Table 14.3 (Continued)

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
P	1	2.01	80.29	0.0120
V	1	6.14	24.84	0.0023
P*V	1	6.14	1.68	0.2414
A	1	6.14	4.18	0.0857
P*A	1	6.14	0.24	0.6413
V*A	1	6.14	0.01	0.9339
P*V*A	1	6.07	0.17	0.6908
TIME	2	18	104.55	<.0001
P*TIME	2	18	0.51	0.6102
V*TIME	2	18	4.62	0.0240
P*V*TIME	2	18	0.70	0.5088
A*TIME	2	18	7.30	0.0048
P*A*TIME	2	18	0.05	0.9486
V*A*TIME	2	18	0.51	0.6102

Least Squares Means

Effect	P	V	A	TIME	Estimate	Standard Error	DF	t Value	Pr > t
P	1				29.4167	1.6868	2.01	17.44	0.0032
P	2				50.7917	1.6868	2.01	30.11	0.0011
V		1			36.5000	1.3948	3.6	26.17	<.0001
V		2			43.7083	1.3948	3.6	31.34	<.0001
A			1		38.6250	1.3948	3.6	27.69	<.0001
A			2		41.5833	1.3948	3.6	29.81	<.0001
TIME				1	37.3750	1.2058	2.1	31.00	0.0008
TIME				2	40.1875	1.2058	2.1	33.33	0.0007
TIME				3	42.7500	1.2058	2.1	35.45	0.0006
V*TIME	1	1			34.0000	1.4170	3.83	23.99	<.0001
V*TIME	1	2			36.8750	1.4170	3.83	26.02	<.0001
V*TIME	1	3			38.6250	1.4170	3.83	27.26	<.0001
V*TIME	2	1			40.7500	1.4170	3.83	28.76	<.0001
V*TIME	2	2			43.5000	1.4170	3.83	30.70	<.0001
V*TIME	2	3			46.8750	1.4170	3.83	33.08	<.0001
A*TIME			1	1	36.3750	1.4170	3.83	25.67	<.0001
A*TIME			1	2	38.2500	1.4170	3.83	26.99	<.0001
A*TIME			1	3	41.2500	1.4170	3.83	29.11	<.0001
A*TIME			2	1	38.3750	1.4170	3.83	27.08	<.0001
A*TIME			2	2	42.1250	1.4170	3.83	29.73	<.0001
A*TIME			2	3	44.2500	1.4170	3.83	31.23	<.0001

- (ii) We use the option “ddfm=satterth” to determine appropriate d.f.
- (iii) The random terms “CH(P)” and “V*A*CH(P)” describe in technical terms the errors $E(A)$ and $E(B)$, respectively, where $V*A*CH(P)$ includes $V*CH(P)$ and $A*CH(P)$.
- (iv) We specify AR(1) as the covariance structure for the repeated measures.
- (v) “PLANT(P*V*A)” specifies the subject for the repeated measures.

The results of the analysis are given in Table 14.3b:

- (vi) The correlation between adjacent observations is obtained as $r = .9355$.
- (vii) The estimates of the variance components $\sigma_{e(A)}^2$, $\sigma_{e(B)}^2$, $\sigma_{e(C)}^2$ are obtained as $CH(P)=3.60$, $V*A*CH(P) = 0$, $Residual = 8.87$, respectively.
- (viii) Since with the inclusion of TIME, that is, repeated measures in the design (see Section 13.6), there will be three different error terms for testing hypotheses about fixed effects: $E(A)$ for P with 2 d.f, $E(B)$ for V, A, $V*A$, $P*V$, $P*A$, $P*V*A$ with 6 d.f. (determined by SAS to be 6.14 or 6.07), and $E(C)$ for all terms involving TIME with 18 d.f. (these include 2 d.f. from the $P*V*A*TIME$ interaction)
- (ix) P, V, A, TIME, $V*TIME$, $A*TIME$ are found to be significant.
- (x) A look at the LS means indicate that the significant interactions are co-directional. Hence, testing hypotheses about main effects is meaningful. \square

14.4 EXERCISES

1. Using the data from Example 14.3 perform the trend analysis as described in Section 14.2.3 and compare the results with those obtained in Table 14.2b.
2. Using the data from Example 14.3 perform the analysis by making comparison at each time point. Compare the results with those obtained in Exercise 1 and those in Table 14.1b.
3. Using the data from Example 14.3 perform the mixed model analysis using the assumption of compound symmetry. Compare the results to those obtained in Table 14.3b.
4. Consider a CRD with subsampling and repeated measures for each subsample. Discuss how you would analyze the data and how you would perform the analysis using SAS PROC MIXED.
5. Consider an RCBD with repeated measures. Discuss how you would analyze the data and how you would perform the analysis using SAS PROC MIXED.

This Page Intentionally Left Blank

Epilogue

Let us now return to the conversation between the statistician (S) and the research scientist (R) described in the Preface.

Several weeks after this conversation R and his research assistant (RA) pay a visit to S who also has one of his graduate students (GS) in his office.

R: "Thank you for taking the time to talk to us. Our paper has been tentatively accepted for publication, but the editor asked, among other things, for some clarification on our experimental design and the analysis. And this is where we hope you can help us. Since RA has done most of the work I'll let her tell you what we did. "

RA: "I made five separate preparations for each of the three types of growth medium. Each preparation contained enough material to fill four pots. On a bench in the greenhouse I arranged 15 rows of 4 pots each. I randomly assigned the 15 preparations to the 15 rows of pots and filled each pot in a row with the assigned growth medium., In each row I then planted one flower from each of the four varieties in a separate pot. The plants were randomized separately in each row."

S: "That is very good. What can you tell me about the environmental conditions in the greenhouse. For example, is the amount of light different at both ends of the bench?"

RA: "I understand what you mean, but the bench is arranged such that the light and temperature conditions are uniform over the entire bench. Also, all the plants were treated identically. For example, they all received the same amount of water, all at the same times. So, there should be no environmental differences."

S: "Fine. Now, for the analysis, what kind of data do you have?"

R: "To evaluate the effectiveness of the growth media we developed an index which combined various aspects of growth, such as height, development of foliage, formation of buds and flowers. We made these observations every two weeks for 12 weeks. For the publication we reported only the results for week 12, because that is most important from an economical point of view, that is, from the producer's point of view. From a scientific point of view it would also be informative to analyze the entire data set over the course of the 12 weeks, that is, using the 6 measurements (indexes) that we have."

S: "I see. GS has just completed a course on experimental design. I'll ask him to help you with the analysis of the data, which will be different from the one you have performed, because the design is different from the one on which you have based your analysis. He can then help you with the interpretation of the results and explain to what extent they may differ from what you presented in the paper. He can also help you with

the analysis of the 12 weeks data.”

GS: “Yes, I know how to do that.”

R and RA: “Thank you very much for your help.”

And you, the reader, are being challenged to consider this as an additional exercise and prepare a report on the type of design used in this experiment and how the analyses of the data should be performed.

Bibliography

- Addelman, S. (1962). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics*, **4**, 21–46.
- Addelman, S., and O. Kempthorne (1961). Orthogonal Main Effect Plans. Aerospace Research Laboratories, Report No. 79.
- Afrarinejad, K. (1983). Balanced repeated measurement designs. *Biometrika*, **70**, 199–204.
- Amaranthus, M.P., M. G. Nair, T. C. Reid, and D. Steinfeld (2005). Improved *Rhizopogon* mycorrhizal colonization and foliar nutrient levels in ponderosa pine and Douglas-fir with Myconate[®]. *J.Sustainable Forestry*, **20**, 1–14.
- Balaam, L. N. (1968). A two-period design with t^2 experimental units. *Biometrics*, **24**, 61–73.
- Bancroft, T. A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. *Biometrics*, **20**, 427–442.
- Bartlett, M. S. (1947). The use of transformations. *Biometrics*, **3**, 39–52.
- Bechhofer, R. E., and A. C. Tamhane (1981). Incomplete block designs for comparing treatments with a control: General theory. *Technometrics*, **23**, 45–57 (Corrigendum **24**, 171).
- Bechhofer, R. E., and A. C. Tamhane (1983). Design of experiments for comparing treatments with a control: Tables of optimal allocations of observations. *Technometrics*, **25**, 87–95.
- Belsley, D. A., E. Kuh, and R. E. Welsch (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
- Beyer, W. H. (ed.) (1991). *CRC Standard Probability and Statistics Tables and Formulae*. Boca Raton, FL: CRC Press, Inc.
- Birch, J. B., and R. H. Myers (1982). Robust analysis of covariance. *Biometrics*, **38**, 699–713.

- Blaisdell, E. A., Jr., and D. Raghavarao (1980). Partially balanced change-over designs based on m-associate class PBIB designs. *J. Royal Statist. Society B*, **42**, 334–338.
- Bose, R. C. (1938). On the application of Galois fields to the problem of the construction of Hyper-Græco-Latin squares. *Sankhyā*, **3**, 323–338.
- Bose, R. C., and K. R. Nair (1939). Partially balanced incomplete block designs. *Sankhyā*, **4**, 337–372.
- Bose, R. C., S. S. Shrikhande, and E. T. Parker (1960). Further results on the construction of mutually orthogonal Latin squares and the falsity of Euler's conjecture. *Canadian Journal of Mathematics*, **12**, 189–203.
- Bowman, K. O. and M. A. Kastenbaum (1975). Sample size requirement: Single and double classification experiments. In *Selected Tables in Mathematical Statistics*, Vol. III (H. L. Harter and I. B. Owen, eds.), 111–232.
- Box, G. E. P. (1952). Multifactor designs of first order. *Biometrika*, **39**, 49–57.
- Box, G. E. P. (1968). Response surfaces. In *The International Encyclopedia of Social Science*. Macmillan and Free Press, 254–259.
- Box, G. E. P., and D. W. Behnken (1960). Some new three-level designs for the study of quantitative variables. *Technometrics*, **2**, 455–475.
- Box, G. E. P., and D. R. Cox (1964). An analysis of transformations. *J. Royal Statist. Society B*, **26**, 211–252.
- Box, G. E. P., and N. R. Draper (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.*, **54**, 622–654.
- Box, G. E. P., and N. R. Draper (1963). The choice of a second order rotatable design. *Biometrika*, **50**, 335–352.
- Box, G. E. P., and N. R. Draper (1975). Robust designs. *Biometrika*, **62**, 347–352.
- Box, G. E. P., and N. R. Draper (1987). *Empirical Model-Building and Response Surfaces*. New York: Wiley.
- Box, G. E. P., and J. S. Hunter (1957). Multi-factor experimental designs for exploring response surfaces. *Ann. Mathem. Statist.*, **28**, 195–241.
- Box, G. E. P., and J. S. Hunter (1961). The 2^{k-p} fractional factorial designs I. *Technometrics*, **3**, 311–351.
- Box, G. E. P., and K. B. Wilson (1951). On the experimental attainment of optimum conditions (with discussion). *J. Royal Statist. Society, B*, **13**, 1–45.
- Bradley, R. A., and R. E. Odeh (1988). A generating algorithm for linear trend-free and nearly trend-free block designs. *Comm. Statist. B-Simulation Comput.*, 1259–1280.

- Bradley, R. A., and C. M. Yeh (1980). Trend-free block designs: Theory. *Ann. Statist.*, **8**, 883–893.
- Buehler, R. J., B. V. Shah, and O. Kempthorne (1964). Methods of parallel tangents. *Chem. Eng. Progress Symp. Series*, **60**, 1–7.
- Calinski, T., and L. C. A. Corsten (1985). Clustering means in ANOVA by simultaneous testing. *Biometrics*, **41**, 39–48.
- Carmer, S. G., and M. R. Swanson (1973). An evaluation of ten pairwise multiple comparison procedures by Monte Carlo Methods. *J. Amer. Statist. Assoc.*, **68**, 66–74.
- Carter, W. H., Jr., J. L. Wampler, and D. M. Stablein (1983). *Regression Analysis of Survival Data in Cancer Chemotherapy*. New York: Marcel Dekker.
- CASS Principal Investigators and Their Associates: Coronary Artery Surgery Study (CASS) (1983a). A randomized trial of coronary artery bypass surgery: survival data. *Circulation*, **68**, 939–950.
- CASS Principal Investigators and Their Associates: Coronary Artery Surgery Study (CASS) (1983b). A randomized trial of coronary artery bypass surgery: quality of life in patients randomly assigned to treatment groups. *Circulation*, **68**, 951–960.
- Clatworthy, W. H. (1973). *Tables of Two-Associate-Class Partially Balanced Designs*. NBS Applied Mathematics Series, 63.
- Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, **13**, 261–281.
- Cochran, W. G., and G. M. Cox (1957). *Experimental Designs* (2nd ed.). New York: Wiley.
- Coleman, D. E. and D. C. Montgomery (1993). A systematic approach to planning for a designed industrial experiment. *Technometrics*, **35**, 1–12.
- Connor, W. S., and S. Young (1961). *Fractional Factorial Designs for Experiments with Factors at Two and Three Levels*. National Bureau of Standards, Applied Mathematics Series 58.
- Conover, W. J., and R. L. Iman (1982). Analysis of covariance using the rank transformation. *Biometrics*, **38**, 715–724.
- Coons, I. (1957). The analysis of covariance as a missing plot technique. *Biometrics*, **13**, 387–405.
- Cornell, J. A. (1975). Some comments on designs for Cox's mixture polynomial. *Technometrics*, **17**, 25–35.
- Cornell, J. A. (2002). *Experiments with Mixtures* (3rd ed.). New York: Wiley.

- Corsten, L. C. A., and A. C. van Eijnsbergen (1972). Multiplicative effects in two-way analysis of variance. *Statistica Neerlandica*, **26**, 61–68.
- Cox, D. R. (1951). Some systematic experimental designs. *Biometrika*, **38**, 312–323.
- Cox, D. R. (1952). Some recent work on systematic experimental designs. *J. Royal Statist. Society, B*, **14**, 211–219.
- Cox, D. R. (1957). The use of a concomitant variable in selecting an experimental design. *Biometrika*, **44**, 150–158 (Correction **44**, 534).
- Cox, D. R. (1982). Randomization and concomitant variables in the design of experiments. In *Essays in Honor of C.R. Rao* (G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, eds.), 197–202. Amsterdam: North-Holland.
- Cox, D. R. (1984). Interaction (with discussion). *Intern. Statist. Rev.*, **52**, 1–32.
- Cox, D. R. (2006). *Principles of Statistical Inference*. Cambridge: Cambridge University Press.
- Cox, D. R., and P. McCullagh (1982). Some aspects of analysis of covariance. *Biometrics*, **38**, 541–561.
- Craske, M. G., A. J. Lang, D. Aikins, and J. L. Mystkowski (2005). Cognitive behavioral therapy for nocturnal panic. *Behavior Therapy*, **36**, 43–54.
- Daniel, C. (1959). Use of the half-normal plots in interpreting factorial two-level experiments. *Technometrics*, **1**, 311–341.
- Dean, C. A., P. P. Cotterill, and R. D. Burton (2006). Early selection in radiata pine. *Silvae Genetica*, **55**, 182–191.
- Dénes, J., and A. I. Keedwell (1974). *Latin Squares and their Applications*. London: Engl. Univ. Press.
- De Palluel, C. (1788). Sur les avantages et l'économie que procurent les racines employées à l'engrais des moutons à l'étable. *Ann. Agric.*, **14**, 133–139.
- Draper, N. R. (1982). Center points in second-order response surface designs. *Technometrics*, **24**, 127–133.
- Duncan, D. B. (1951). A significance test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, **2**, 171–189.
- Duncan, D. B. (1955). Multiple range and multiple *F*-tests. *Biometrics*, **11**, 1–42.
- Dunlap, W. P., R. S. Powell, and T. K. Konnerth (1977). A FORTRAN IV function for calculating probabilities associated with the studentized range statistic. *Behavior Research Methods and Instrumentation*, **9**, 373–375.
- Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Stat. Assoc.*, **50**, 1096–1121.

- Dunnett, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics*, **20**, 482–491.
- Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, **3**, 1–21.
- Euler, L. (1782). Recherches sur une nouvelle espèce de quarrés magiques. *Verh. Zeeuwach Genootschap Wetenschappen Vlissengen*, **9**, 85–239.
- Everitt, B. S. (1995). The analysis of repeated measures: a practical review with examples. *The Statistician*, **44**, 113–135.
- Federer, W. T. (1984). Principles of statistical design with special reference to experiment and treatment design. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds.). Ames, IA: Iowa State University Press, 77–104.
- Finney, D. J. (1990). Repeated measurements: What is measured and what repeats? *Statistic. Med.*, **9**, 639–644.
- Fisher, R. A. (1921). Studies in crop variation. *J. Agric. Sci.*, **11**, 107–135.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1926). The arrangement of field experiments. *J. Ministry Agric. Engl.*, **33**, 503–513.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh and London: Oliver and Boyd.
- Fisher, R. A. (1937). *The Design of Experiments* (2nd ed.). Edinburgh: Oliver and Boyd.
- Fisher, R. A., and F. Yates (1957). *Statistical Tables for Biological, Agricultural and Medical Research* (5th ed.). New York: Hafner Publ. Co.
- Freeman, M. F., and J. W. Tukey (1950). Transformations related to the angular and the square root. *Ann. Mathem. Statist.*, **21**, 607–611.
- Frison, L. and S. J. Pocock (1990). Repeated measures in clinical trials: analysis using mean summary statistics and its implication for design. *Statist. Med.*, **11**, 1685–1704.
- Gallie, W. B. (1966). *Peirce and Pragmatism*. New York: Dover Publications.
- Geisser, S., and S. Greenhouse (1958). An extension of Box's results on the use of the *F*-distribution in multivariate analysis. *Ann. Mathem. Statist.*, **29**, 885–891.
- Glenn, W. A., and C. Y. Kramer (1958). Analysis of variance of a randomised block design with missing observations. *Appl. Stat.*, **7**, 173–185.
- Hand, D. and M. Crowder (1996). *Practical Longitudinal Data Analysis*. London: Chapman and Hall.

- Harter, H. L. (1960). Critical values for Duncan's new multiple range test. *Biometrics*, **20**, 482–491.
- Hartley, H. O. (1950). The maximum F -ratio as a short cut test for heterogeneity of variance. *Biometrika*, **37**, 308–312.
- Hartley, H. O., and C. A. B. Smith (1948). The construction of Youden squares. *J. Royal Stat. Soc., B*, **10**, 262–263.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.
- Hedayat, A., and K. Afsarinejad (1975). Repeated measurement designs, I. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.), 229–242. Amsterdam: North Holland.
- Hedayat, A., and W. T. Federer (1974). Pairwise and variance balanced incomplete block designs. *Ann. Inst. Stat. Math.*, **26**, 331–338.
- Hering, F. and S. Mejza (1997). Incomplete split-block designs. *Biometrical J.*, **39**, 227–238.
- Hinkelmann, K. (1963). A commonly occurring incomplete multiple classification model. *Biometrics*, **19**, 105–117.
- Hinkelmann, K. (1968). Missing values in partial diallel experiments. *Biometrics*, **24**, 903–913.
- Hinkelmann, K. (2001). Remembering Oscar Kempthorne (1919–2000). *Statistical Science*, **16**, 169–183.
- Hinkelmann, K. (2004). Evaluating and interpreting interaction. Technical Report **04-6**, Department of Statistics, Virginia Tech. (http://www.stat.org.vt/dept/web-e/tech_reports/TechReport04 - 6 .pdf)
- Hinkelmann, K. and J. S. Alcorn (1998). Randomization analysis of replicated complete block designs. *J. Combinatorics, Information and System Science*, **23**, 317–332.
- Hochberg, Y., and A. C. Tamhane (1987). *Multiple Comparison Procedures*. New York: Wiley.
- Hocking, R. R. (1982). Discussion of “Some aspects of analysis of covariance” by D. R. Cox and P. McCullagh. *Biometrics*, **38**, 559–561.
- Hocking, R. R. (1985). *The Analysis of Linear Models*. Monterey, CA: Brooks/Cole.
- Hocking, R. R. (2003). *Methods and Applications of Linear Models* (2nd ed). Hoboken, NJ: Wiley.
- Hocking, R. R. and F. M. Speed (1975). A full rank analysis of some linear model problems. *J. Amer. Statist. Assoc.*, **70**, 706–712.

- Hryniewicz, K., A. S. Androne, A. Hudaihed, and S. D. Katz (2003). Comparative effects of carvedilol and metoprolol on regional vascular responses to adrenergic stimuli in normal subjects and patients with chronic heart failure. *Circulation.*, **108**, 971–976.
- Hsu, J. C. (1996). *Multiple Comparisons—Theory and Methods*. London: Chapman and Hall.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Mathem. Statist.*, **35**, 73–101.
- Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.
- Huynh, H., and L. S. Feldt (1970). Conditions under which mean square ratios in repeated measurements designs have exact *F*-distributions. *J. Amer. Stat. Assoc.*, **65**, 1582–1589.
- Jacroux, M. (1998). On the determinaton and construction of *E*-optimal block designs in the presence of linear trends. *J. Statist. Planning and Inference*, **67**, 331–342.
- Jacroux, M., D. Majumdar, and R. K. Shah (1995). Efficient block designs in the presence of trends. *Statistica Sinica*, **5**, 605–615.
- Jo, J. and K. Hinkelmann (1993). Some properties of Box-Behnken designs. *J. Combinatorics, Information and System Sci.*, **18**, 273–287.
- John, J. A., and M. H. Quenouille (1977). *Experiments: Design and Analysis*. London: Griffin.
- John, P. W. M. (1964). Balanced designs with unequal numbers of replicates. *Ann. Mathem. Statist.*, **35**, 897–899.
- Johnson, D. E., and F. A. Graybill (1972). An analysis of a two-way model with interaction and no replication. *J. Amer. Stat. Assoc.*, **67**, 862–868.
- Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions*, Volume 2 (2nd ed). New York: Wiley.
- Jones, B. and M. G. Kenward (2003). *Design and Analysis of Cross-over Trials*, (2nd ed). Boca Raton, FL: Chapman and Hall/CRC.
- Kabelka, E. A., S. R. Carlson, and B. W. Diers. (2006). *Glycine soja* PI 468916 SCN resistance loci's associated effects on soybean seed yield and other agronomic traits. *Crop Science*, **46**, 622–629.
- Kempthorne, O. (1952). *Design and Analysis of Experiments*. New York: Wiley.
- Kempthorne, O. (1955). The randomization theory of experimental inference. *J. Amer. Statist. Assoc.*, **50**, 946–967.

- Kempthorne, O. (1971). Probability, statistics, and the knowledge business. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Toronto: Holt, Rinehart and Winston, 470–499.
- Kempthorne, O. (1972). Theories of inference and data analysis. In *Statistical Papers in Honor of George W. Snedecor* (T. A. Bancroft, ed.). Ames, IA: Iowa State University Press, 167–191.
- Kempthorne, O. (1973). Statistics and the philosophers. In *Foundations and Philosophy of Statistical Theories in the Physical Sciences*, Vol. II, 273–314. London, Ontario: Univ. of Western Ontario.
- Kempthorne, O. (1975). Inference from experiments and randomization. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.). Amsterdam: North-Holland, 303–331.
- Kempthorne, O. (1976). Best linear unbiased estimation with arbitrary variance matrix. In *Essays in Probability and Statistics* (S. Ikeda et al., eds.), 203–225.
- Kempthorne, O. (1992). Intervention experiments, randomization and inference. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*. (M. Ghosh and P. K. Pathak, eds.). Inst. of Mathem. Statistics Lecture Notes—Monograph Series, Vol. 17, 13–31.
- Kempthorne, O., and T. E. Doerfler (1969). The behaviour of some significance tests under experimental randomization. *Biometrika*, **56**, 231–248.
- Kempthorne, O., and L. Folks (1971). *Probability, Statistics, and Data Analysis*. Ames, IA: Iowa State University Press.
- Kempthorne, O., G. Zyskind, S. Addelman, T. N. Throckmorton, and R. F. White (1961). *Analysis of Variance Procedures*. Aeronautical Research Laboratory Techn. Report 149, Wright-Patterson Air Force Base, Ohio.
- Keppel, G., and S. Zedeck (1989). *Data Analysis for Research Designs*. New York: Freeman.
- Keselman, H. J., J. Algina, and R. K. Kowalchuk (2001). The analysis of repeated measures designs: A review. *British J. Mathem. Statist. Psychology*, **54**, 1–20.
- Keynes, J. M. (1921). *A Treatise on Probability*. New York: Macmillan.
- Khuri, A. I., and J. A. Cornell (1996). *Response Surfaces: Designs and Analyses* (2nd ed). New York: Marcel Dekker.
- Kiefer, J. (1959). Optimum experimental design. *J. Royal Stat. Soc., B*, **21**, 272–319.
- Kii, W. Y. and G. McL. Dryden (2005). Effect of drinking saline water on food and water intake, food digestibility, and nitrogen mineral balances of rusa deer stags (*Cervus timorensis russa*). *Animal Science*, **81**, 99–105.

- Kirk, R. F. (1982). *Experimental Design: Procedures for the Behavioral Sciences* (2nd ed). Belmont, CA: Wadsworth.
- Kramer, C. Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, **12**, 307–310.
- Kramer, C. Y. (1957). Extension of multiple range tests to group correlated adjusted means. *Biometrics*, **13**, 13–18.
- Kramer, C. Y., and S. Glass (1960). Analysis of variance of a Latin square design with missing observations. *Appl. Stat.*, **9**, 43–50.
- Kress, L. W., J. M. Skelly, and K. Hinkelmann (1982a). Growth impact of O₃, NO₂ and/or SO₂ on *Platanus Occidentalis*. *Agric. and Environment*, **7**, 265–274.
- Kress, L. W., J. M. Skelly, and K. Hinkelmann (1982b). Relative sensitivity of 18 full-sib families of *Pinus taeda* to O₃. *Can. J. Forest Res.*, **12**, 203–209.
- Kuehl, R. O. (1994). *Statistical Principles of Research Design and Analysis*. Belmont, CA: Duxbury.
- Lane, P. W., and J. A. Nelder (1982). Analysis of covariance and standardization as instances of prediction. *Biometrics*, **38**, 613–621.
- Lencina, V. B., J. M. Singer, and E. J. Stanek, III (2005). Much ado about nothing: the mixed models controversy revisited. *Intern. Statist. Review*, **73**, 9–20.
- Lentner, M., J. C. Arnold, and K. Hinkelmann (1989). The efficiency of blocking: How to use MS(Blocks)/MS(Error) correctly. *Amer. Statistician*, **43**, 106–108.
- Lucas, H. L. (1957). Extra-period Latin square change-over designs. *J. Dairy Science*, **40**, 225–239.
- Mandel, J. (1961). Non-additivity in two-way analysis of variance. *J. Amer. Statist. Assoc.*, **56**, 878–888.
- Mandel, J. (1971). The new analysis of variance model for non-additive data. *Technometrics*, **13**, 1–18.
- Marasinghe, M. G. (1985). Asymptotic tests and Monte Carlo studies associated with multiplicative interaction model. *Comm. Stat. A.*, **14**, 2219–2231.
- Mathon, R. and A. Rosa (1996). $2 - (v, k, \lambda)$ designs of small order. In: *The CRC Handbook of Combinatorial Designs*, (C. J. Colbourn and J. H. Dinitz, eds.). Boca Raton, FL: CRC, 3–41.
- Matthews, J. N. S., D. G. Altman, M. J. Campbell, and P. Royston (1990). Analysis of serial measurements in medical research. *Br. Med. J.*, **300**, 230–235.
- McCarthy, M. D. (1937). On the application of the z -test to randomized blocks. *Annals of Mathematical Statistics*, **10**, 337–359.

- McLean, R. A., and V. L. Anderson (1984). *Applied Factorial and Fractional Designs*. New York: Marcel Dekker.
- Mead, R., and D. J. Pike (1975). A review of response surface methodology from a biometric viewpoint. *Biometrics*, **31**, 803–851.
- Miller, R. G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Springer.
- Milliken, G. A., and D. E. Johnson (1984). *Analysis of Messy Data. Vol. I: Designed Experiments*. Belmont, CA: Lifetime Learning Publications.
- Myers, R. H. (1990). *Classical and Modern Regression with Applications* (2nd ed.). Boston: PWS-Kent Publ. Co.
- Myers, R. H. and D. C. Montgomery (2002). *Response Surface Methodology: Process and Product Optimization using Designed Experiments* (2nd ed.). New York: Wiley.
- Narula, S. C. (1978). Orthogonal polynomial regression for unequal spacing and frequencies. *J. Quality Technology*, **10**, 170–179.
- Nelder, J. A. (1994). The statistics of linear models: back to basics. *Statistics and Computing*, **4**, 221–234.
- Nelder, J. A. (1995). Rejoinder to comments on “The statistics of linear models: back to basics”. *Statistics and Computing*, **5**, 109–111.
- Neyman, J. (1929). *The Theoretical Basis of Different Methods of Testing Cereals. Part I*. Warsaw: K. Buszczyński and Sons.
- Neyman, J. (1950). *First Course in Probability and Statistics*. New York: Holt.
- Neyman, J. (with K. Iwazkiewicz and S. Kolodziejczyk) (1935). Statistical problems in agricultural experimentation. *J. Royal Stat. Soc., B*, **2**, 107–180.
- Northrop, F. S. C. (1947). *Logic of Sciences and the Humanities*. New York: Macmillan.
- Nowell-Smith, P. H. (1960). Causality or Causation. In *Encyclopædia Britannica*, Vol. 5. London: Encyclopædia Britannica Co.
- Odeh, R. E., and M. Fox (1975). *Sample Size Choice: Charts for Experiments with Linear Models*. New York: Marcel Dekker.
- Parker, P. A., S. M. Kowalski, and G. G. Vining (2007). Construction of balanced equivalent estimation second-order split-plot designs. *Technometrics*, **49**, 56–65.
- Patterson, H. D. (1950). The analysis of change-over trials. *J. Agric. Sci.*, **40**, 375–380.
- Patterson, H. D. (1951). Change-over trials. *J. Royal Statist. Soc., B*, **13**, 256–271.
- Patterson, H. D. (1952). The construction of balanced designs for experiments involving sequences of treatments. *Biometrika*, **39**, 32–48.

- Patterson, H. D., and H. L. Lucas (1959). Extra-period change-over designs. *Biometrics*, **15**, 116–132.
- Pauling, L. (1970). *Vitamin C and the Common Cold*. San Francisco, CA: Freeman.
- Pearce, S. C. (1953). Field experimentation with fruit trees and other perennial plants. *Commonwealth Agric. Bus. Techn. Communic.*, **23**.
- Pearce, S. C. (1960). Supplemental balance. *Biometrika*, **47**, 263–271 (Corrigenda, **48**, 475).
- Pearce, S. C. (1983). *The Agricultural Field Experiment: A Statistical Examination of Theory and Practice*. New York: Wiley.
- Pearson, E. S., and H. O. Hartley (1970). *Biometrika Tables for Statisticians*, Vol. I. Cambridge: Cambridge Univ. Press.
- Peirce, C. S. (1966). *Collected papers, Vol. 5: Pragmatism and Pragmaticism* (C. Hartshorne and P. Weiss, eds). Cambridge, MA: Harvard University Press.
- Perry, C. O. (1983). Diallel analysis of rind puncture and grain yield and their interactions with plant densities for twelve elite inbred lines of maize, Zea Mays, L. Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.
- Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations: III. The analysis of variance test. *Biometrika*, **29**, 322–335.
- Plackett, R. L., and J. P. Burman (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–325.
- Preece, D. A. (1982). Balance and designs: Another terminological tangle. *Utilitas Mathematica*, **21C**, 85–186.
- Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments*. New York: Wiley.
- Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. New York: Wiley.
- Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. on Mathem. Statist. and Probability I*, 355–372, Univ. of California Press.
- Rao, C. R. (1971). Unified theory of linear estimation. *Sankhyā A*, **33**, 370–396 (and *Sankhyā A*, **34**, 477).
- Rao, C. R. (1973). Representations of best linear unbiased estimators in the Gauss-Markoff model with singular dispersion matrix. *J. Multivariate Analysis*, **3**, 276–292.
- Ratkowsky, D. A., M. A. Evans, and J.R. Alldredge (1993). *Cross-over Experiments—Design, Analysis and Applications*. New York: Dekker.

- Richards, W. (1980). The Randomization Analysis of Covariance and change-over designs. Unpublished Ph.D. dissertation, Iowa State Univ., Ames, IA.
- Ringland, J. T. (1983). Robust multiple comparisons. *J. Amer. Stat. Assoc.*, **78**, 145–151.
- Robinson, J. (1967). Incomplete split-plot designs. *Biometrics*, **23**, 793–802.
- Robinson, J. (1973). The analysis of covariance under randomization model. *J. Royal Stat. Soc., B*, **35**, 368–376.
- Robinson, J. (1975). On the test for additivity in a randomized block design. *J. Amer. Stat. Assoc.*, **70**, 184–185.
- Rogers, W. J., C. J. Coggin, B. J. Gersh, L. D. Fisher, W. D. Myers, A. Oberman, and L. T. Sheffield (1990). Ten-year follow-up of quality of life in patients randomized to receive medical therapy or coronary artery by-pass graft surgery. *Circulation*, **82**, 1647–1658.
- Rojas, B. A. (1973). On Tukey's test of additivity. *Biometrics*, **29**, 45–52.
- Rosen, C. J., V. A. Fritz, G. M. Gardner, S. S. Hecht, S. G. Carmella, and P. M. Kenney (2005). Cabbage yield and glucosinolate concentrations as affected by nitrogen and sulfur fertility. *HortScience*, **40**, 1493–1498.
- Rosenbaum, P. R. (1984). Conditional permutation tests and the propensity score in observational studies. *J. Amer. Statist. Assoc.*, **79**, 41–48.
- Roux, C. Z. (1984). Treatment by unit interactions in the completely randomized and randomized block designs. In *Experimental Design, Statistical Models and Genetic Statistics* (K. Hinkelmann, ed.), 141–153. New York: Marcel Dekker.
- Rowell, J. G., and D. E. Walters (1976). Analyzing data with repeated observations on each experimental unit. *J. Agric. Sci. Cambridge*, **87**, 423–432.
- Roy, R. K. (1990). *A Primer on the Taguchi Method*. New York: Van Nostrand Reinhold.
- Runes, D. D. (1962). *Dictionary of Philosophy*. Totawa, NJ: Littlefield, Adams and Company.
- Russell, B. (1959). *Wisdom of the West*. New York: Crescent Books.
- SAS Institute, Inc. (2002–2003). *SAS/STATTM User's Guide, Release 9.1 Edition*. Cary, NC: SAS Inst.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2**, 110–114.
- Savage, L. J. (1972). *The Foundations of Statistics* (2nd ed.). New York: Dover.

- Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87–104.
- Scheffé, H. (1958). Experiments with mixtures. *J. Royal Stat. Soc., B*, **20**, 344–360.
- Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.
- Scheffler, Isreal (1967). *Science and Subjectivity*. Indianapolis: Bobbs-Merrill.
- Scott, A. J., and M. Knott (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, **30**, 507–512.
- Searle, S. R., F. M. Speed, and G. A. Milliken (1980). Population marginal means in the linear model: An alternative to least squares means. *Amer. Statistician*, **34**, 216–221.
- Shah, B. V. (1958). On balancing in factorial experiments. *Ann. Mathem. Statist.*, **29**, 766–779.
- Shah, B. V., R. J. Buehler, and O. Kempthorne (1964). Some algorithms for minimizing a function of several variables. *J. Soc. Indust. Appl. Math.*, **12**, 74–92.
- Smith, H. F. (1957). Interpretation of adjusted treatment means and regressions in analysis of covariance. *Biometrics*, **13**, 282–308.
- Snedecor, G. W. (1946). *Statistical Methods*. Ames, IA: Iowa State College Press.
- Snee, R. D. (1982). Nonadditivity in a two-way classification: Is it interaction or non-homogeneous variance? *J. Amer. Statist. Assoc.*, **77**, 515–519.
- Spedding, J., R. L. Ellis, and D. D. Heath (1861). *The Works of Francis Bacon*, Vol. I. Boston: Brown and Taggard.
- Spedding, J., R. L. Ellis, and D. D. Heath (1863). *The Works of Francis Bacon*, Vol. VIII. Boston: Taggard and Thompson.
- Spendley, W., G. R. Hext, and F. R. Hinsworth (1962). Sequential application of simplex designs in optimization and EVOP. *Technometrics*, **4**, 441–461.
- Srivastava, J. (1984). Sensitivity and revealing power: Two fundamental statistical criteria other than optimality arising in discrete experimentation. In *Experimental Design, Statistical Models, and Genetic Statistics* (K. Hinkelmann, ed.). New York: Marcel Dekker, 95–117.
- Srivastava, J. N. and Wang, Y. C. (1998). Row-column designs: Non-additivity makes them hazardous to use. *Journal of Statistical Planning and Inference*, **73**, 277–315.
- Stebbing, L. S. (1937). *Philosophy and the Physicist*. Middlesex, England: Penguin Books.
- Stevens, W. L. (1938). The completely orthogonalized square. *Annals of Eugenics*, **9**, 83–93.

- Stewart, M. J. (1980). Randomization analysis of replicated randomized complete block designs. Unpublished Ph.D. thesis, Virginia Polytechnic Institute and State Univ., Blacksburg, VA.
- Stoline, M. R. (1981). The status of multiple comparisons: Simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. *Amer. Statist.*, **35**, 134–141.
- Street, A. P., and D. J. Street (1987). *Combinatorics of Experimental Design*. Oxford: Clarendon Press.
- Street, A. P., and D. J. Street (1988). Latin squares and agriculture: The other bicentennial. *Math. Scientist*, **13**, 48–55.
- Stufken, J. (1988). On the existence of linear trend-free block designs. *Commun. Statist-Theory Meth.*, **17**, 3857–3863.
- Taguchi, G. (1986). *Introduction to Quality Engineering*. Asian Productivity Organization, Dearborn, MI: American Supplier Inst.
- Taguchi, G. (1987). *Systems of Experimental Design*. New York: UNIPUB/Kraus International Publications.
- Tang, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. *Stat. Research Memoirs*, **2**, 126–149.
- Throckmorton, T. N. (1961). Structures of Classification Data. Unpublished Ph.D. Dissertation, Iowa State University, Ames, IA.
- Tukey, J. W. (1949). One degree of freedom for nonadditivity. *Biometrics*, **5**, 232–242.
- Tukey, J. W. (1955). Query 113. *Biometrics*, **11**, 111–113.
- Tukey, J. W. (1977). Some thoughts on clinical trials, especially problems of multiplicity. *Science*, **198**, 679–684.
- van Belle, G., L. D. Fisher, P. J. Heagerty and T. Lumley (2004). *Biostatistics—A Methodology for the Health Sciences* (2nd ed.). Hoboken, NJ: Wiley.
- Venn, J. (1962). *The Logic of Chance* (4th ed.). New York: Chelsea.
- Vining, G. G., and R. H. Myers (1990). Combining Taguchi and response surface philosophies: A dual response approach. *J. Quality Technology*, **22**, 38–45.
- Voss D. T. (1999). Resolving the mixed model controversy. *Amer. Statist.*, **53**, 352–356.
- Wagenaar, W. A. (1969). A note on the construction of diagram-balanced Latin squares. *Psych. Bull.*, **72**, 384–386.
- Ward, G. C., and I. D. Dick (1952). Non-additivity in randomized block designs and balanced incomplete block designs. *New Zealand J. Sci. Tech., B*, **33**, 430–435.

- Watson, G. S. (1967). Linear least squares regression. *Ann Mathem. Statist.*, **38**, 1679–1699.
- Webb, D. F., J. M. Lucas, and J. J. Borkowski (2004). Factorial experiments when factor levels are not necessarily reset. *J. Quality Technology*, **36**, 1–11.
- Welch, B. L. (1937). On the z -test in randomized blocks and Latin squares. *Biometrika*, **29**, 21–52.
- Westfall, P. H., R. D. Tobias, D. Rom, R. D. Wolfinger, and Y. Hochberg (1999). *Multiple Comparisons and Multiple Test—Using the SAS® System*. Cary, NC: SAS Institute Inc.
- Wilk, M. B. (1955). The randomization analysis of a generalized randomized block design. *Biometrika*, **42**, 70–79.
- Wilk, M. B., and O. Kempthorne (1955). Fixed, mixed and random models. *J. Amer. Stat. Assoc.*, **50**, 1144–1167.
- Wilk, M. B., and O. Kempthorne (1956). Some aspects of the analysis of factorial experiments in a completely randomized design. *Ann. Math. Stat.*, **27**, 950–985.
- Wilk, M. B., and O. Kempthorne (1957). Non-additivities in a Latin square design. *J. Amer. Stat. Assoc.*, **52**, 218–236.
- Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian J. Sci. Research, A*, **2**, 149–168.
- Williams, E. J. (1950). Experimental designs balanced for pairs of residual effects. *Australian J. Sci. Research, A*, **3**, 351–363.
- Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Emp. J. Exp. Agr.*, **1**, 129–142.
- Yates, F. (1935). Complex experiments. *J. Roy. Statist. Soc., Suppl.* **2**, 181–247.
- Yates, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.*, **26**, 424–455.
- Yates, F. (1937). *The Design and Analysis of Factorial Experiments*. Commonwealth Bureau of Soil Science, Tech. Comm., No. 35.
- Yates, F. (1937). *The Design and Analysis of Factorial Experiments*. Harpenden: Imp. Bur Soil Sci.
- Yeh, C. M., and R. A. Bradley (1983). Trend-free block designs: Existence and construction results. *Commun. Statist.-Theory Meth.*, **12**, 1–24.
- Yeh, C. M., R. A. Bradley, and W. I. Notz (1985). Nearly trend-free block designs. *J. Amer. Statist. Assoc.*, **80**, 985–992.

- Youden, W. J. (1937). Use of incomplete block replications in estimating tobacco-mosaic virus. *Contr. Boyce Thompson Inst.*, **9**, 41–48.
- Zahn, D. A. (1975). Modifications of revised critical values for the half-normal plot. *Technometrics*, **17**, 189–200.
- Zyskind, G. (1962). On structure, relation, Σ , and expectation of mean squares. *Sankhyā, A*, **24**, 115–148.
- Zyskind, G. (1967). On canonical forms, negative covariance matrices and best and simple least squares estimator in linear models. *Ann. Mathem. Statist.*, **38**, 1092–1100.

Abbreviations

ANOVA - analysis of variance	LSD - Latin square design
AR(1) - first order autocorrelation	LSM - least squares mean
BIBD - balanced incomplete block design	MS - mean square
BLUE - best linear unbiased estimator	NE - normal equations
BTIBD - balanced treatment incomplete block design	OLS - ordinary least squares
CCD - central composite design	OR - operational region
CRD - completely randomized design	OU - observational unit
CS - compound symmetry	PARTAN - parallel tangents method
CWE - comparisonwise error rate	PBIBD - partially balanced incomplete block design
d.f. - degrees of freedom	PE - pure error
ER - experimental region	PFE - per family error rate
ERE - estimated relative efficiency	PMM - population marginal mean
ETC - easy-to-change factor	RCBD - randomized complete block design
EU - experimental unit	RE - relative efficiency
FWE - familywise error rate	REML - residual (restricted) maximum likelihood
GLS - generalized least squares	RHS - right-hand side
GMLM - Gauss-Markov linear model	RNE - reduced normal equations
GMNLM - Gauss-Markov normal linear model	RSM - response surface methods
GRBD - generalized randomized block design	se - standard error
HTC - hard-to-change factor	s.i.p. - symmetric and idempotent
IBD - incomplete block design	SPD - split-plot design
i.i.d. - independently identically dis- tributed	SP(POW) - spatial power
IMSE - integrated mean squared error	SS - sum of squares
LOF - lack of fit	SSPD - split-split-plot design
	UN - unstructured

This Page Intentionally Left Blank

Author Index

- Addelman, S., 109, 480
Afsarinejad, K., 401-402, 573
Aikins, D., 33
Alcorn, J. S., 308
Algina, J., 574
Alldredge, J. R., 407
Altman, D. G., 574
Amaranthus, M. P., 33
Anderson, V. L., 479
Androne, A. S., 32
Arnold, J. C., 292, 389
- Balaam, L. N., 399, 403
Balakrishnan, N., 175
Bancroft, T. A., 256, 313
Bartlett, M. S., 198
Bechhofer, R. E., 333, 335
Behnken, D. W., 509, 511
Belsley, D. A., 258
Beyer, W. H., 221
Birch, J. B., 257
Blaisdell, E. A., Jr., 403
Borkowski, J. J., 511
Bose, R. C., 335, 396-397
Bowman, K. O., 184, 190, 196, 292
Box, G. E. P., 200, 455, 497, 499, 504, 506-507, 509, 517-519
Bradley, R. A., 341-342
Buehler, R. J., 519
Burman, J. P., 464
Burton, R. D., 306
- Calinski, T., 229-230
Campbell, M. J., 575
Carlson, S. R., 307
- Carmella, S. G., 38
Carmer, S. G., 225
Carter, W. H., Jr., 497
CASS Principal Investigators, 20
Clatworthy, W. H., 335-336, 509
Cochran, W. G., 239, 253, 331, 395-396, 400, 542
Coggin, C. J., 20
Coleman, D. E., 30
Connor, W. J., 479
Conover, W. J., 258
Coons, I., 295, 389
Cornell, J. A., 499, 504, 507, 517-519, 521-522
Corsten, L. C. A., 229-230, 306
Cotterill, P. P., 306
Cox, D. R., 35, 151, 200, 252-253, 260, 292, 341
Cox, G. M., 331, 395-396, 400, 542
Craske, M. G., 33
Crowder, M., 573
- Daniel, C., 441
Dean, C. A., 306
Dénes, J., 376
De Palluel, C., 374
Dick, I. D., 303
Diers, B. W., 307
Doerfler, T. E., 127, 178, 286
Draper, N. R., 499, 507, 509, 517-518
Dryden, G. McL., 33
Duncan, D. B., 226
Dunlap, W. P., 230
Dunnett, C. W., 227-228

- Eisenhart, C., 132, 326
 Ellis, R. L., 3, 7
 Euler, L., 374
 Evans, M. A., 407
 Everitt, B. S., 574

 Federer, W. T., 59, 338
 Feldt, L. S., 545, 577
 Finney, D. J., 574, 579
 Fisher, L. D., 20, 313, 319
 Fisher, R. A., 68, 139-140, 148-150, 177,
 221, 225, 239, 278, 286, 374,
 377, 396, 420
 Folks, L., 11, 27
 Fox, M., 184
 Freeman, M. F., 200
 Frison, L., 573, 575
 Fritz, V. A., 38

 Gallie, W. B., 7
 Gardner, G. M., 38
 Geisser, S., 577
 Gersh, B. J., 20
 Glass, S., 389
 Glenn, W. A., 298
 Graybill, F. A., 305-306
 Greenhouse, S., 577

 Hand, D., 573
 Harter, H. L., 226-227
 Hartley, H. O., 184, 221, 323, 395
 Harville, D. A., 130
 Heagerty, P. J., 20, 313, 319
 Heath, D. D., 3, 7
 Hecht, S. S., 38
 Hedayat, A., 338, 401, 573
 Hering, F., 560
 Hext, G. R., 519
 Hinkelmann, K., xix, 35, 38, 41, 55, 292,
 295, 308, 389, 428, 509
 Hinsworth, F. R., 519
 Hochberg, Y., 225-226, 228, 252
 Hocking, R. R., 99-100, 134, 258
 Hryniewicz, K., 32
 Hsu, J. C., 225
 Huber, P. J., 228, 257

 Hudaihed, A., 32
 Hunter, J. S., 455, 499, 506-507
 Huynh, H., 545, 577

 Iman, R. L., 258

 Jacroux, M., 342
 Jo, J., 509
 John, J. A., 366
 John, P. W. M., 338
 Johnson, D. E., 305-306, 570-571
 Johnson, N. L., 175
 Jones, B., 400

 Kabelka, E. A., 307
 Kastenbaum, M. A., 184, 190, 196, 292
 Katz, S. D., 32
 Keedwell, A. I., 376
 Kempthorne, O., xxii, 11, 27, 109, 127,
 151, 157, 161, 174, 178, 197-
 198, 228, 285-286, 291, 300,
 312, 323-324, 377, 386-387,
 419, 468, 480, 519
 Kenney, P. M., 38
 Kenward, M. G., 400
 Keppel, G., 548
 Keselman, H. J., 574
 Keynes, J. M., 25
 Khuri, A. I., 499, 504, 507, 517-518
 Kiefer, J., 59
 Kii, W. Y., 33
 Kirk, R. E., 307
 Knott, M., 229
 Konnerth, T. K., 230
 Kotz, S., 175
 Kowalchuk, R. K., 574
 Kowalski, S. M., 512
 Kramer, C. Y., 226, 251-252, 298, 389
 Kress, L. W., 41
 Kuehl, R. O., 232
 Kuh, E., 258

 Lane, P. W., 245
 Lang, A. J., 33
 Lencina, V. B., 134
 Lentner, M., 292, 389

- Lucas, H. L., 400, 403
 Lucas, J. M., 511
 Lumley, T., 20, 313, 319

 Majumdar, D., 342
 Mandel, J., 302-303, 305-306, 440
 Marasinghe, M. G., 306
 Mathon, R., 331
 Matthews, J. N. S., 575
 McCarthy, M. D., 197
 McCullagh, P., 253, 260
 McLean, R. A., 479
 Mead, R., 497
 Mejza, S., 560
 Miller, R. G., 225-228, 251
 Milliken, G. A., 245, 570-571
 Montgomery, D. E., 30, 499, 507
 Myers, R. H., 74, 257-258, 263, 498-499, 507
 Myers, W. D., 20
 Mystkowski, J. L., 33

 Nair, K. R., 335
 Nair, M. G., 33
 Narula, S. C., 224
 Nelder, J. A., 134, 245
 Neyman, J., 140, 197, 341, 387
 Northrop, F. S. C., 24
 Notz, W. I., 342
 Nowell-Smith, P. H., 18

 Oberman, A., 20
 Odeh, R. E., 184, 342

 Parker, E. T., 396
 Parker, P. A., 512
 Patterson, H. D., 402-403
 Pauling, L., 14
 Pearce, S. C., 35, 277, 333, 428
 Pearson, E. S., 184, 221
 Peirce, C. S., 19
 Perry, C. O., 295
 Pike, D. J., 497
 Pitman, E. J. G., 177, 286
 Plackett, R. L., 464
 Pocock, S. J., 573, 575

 Powell, R. S., 230
 Preece, D. A., 59

 Quenouille, M. H., 366

 Raghavarao, D., 331, 403
 Rao, C. R., 127
 Ratkowsky, D. A., 407
 Reid, T. C., 33
 Richards, W., 176
 Ringland, J. T., 228
 Robinson, J., 248, 305, 549-550
 Rogers, W. J., 20
 Rojas, B. A., 387
 Rom, D., 225
 Rosa, A., 331
 Rosen, C. J., 38
 Rosenbaum, P. R., 240
 Roux, C. Z., 303
 Rowell, J. G., 575
 Roy, R. K., 477
 Royston, P., 575
 Runes, D. D., 17
 Russell, B., 7, 12-13

 SAS Institute, Inc., xviii, 69, 86, 98, 154, 180, 185, 201, 343, 451, 459, 578
 Satterthwaite, F. E., 326, 542, 561
 Savage, L. J., 26
 Scheffé, H., 184, 228, 260, 303, 305, 388, 519
 Scheffler, I., 24
 Scott, A. J., 229
 Searle, S. F., 245
 Shah, B. V., 59, 519
 Shah, R. K., 342
 Sheffield, L. T., 20
 Shrikhande, S. S., 396
 Singer, J. M., 134
 Skelly, J. M., 41
 Smith, C. A. B., 399
 Smith, H. F., 240, 256
 Snedecor, G. W., 229
 Snee, R. D., 312
 Spedding, J., 3, 7

- Speed, F. M., 100, 245
 Spendley, W., 519
 Srivastava, J., 60, 387
 Stablein, D. M., 497
 Stanek, E. J., III, 134
 Stebbing, L. S., 21
 Steinfeld, D., 33
 Stevens, W. L., 397
 Stewart, M. J., 308
 Stoline, M. R., 225
 Street, A. P., 374
 Street, D. J., 374
 Stufken, J., 342
 Swanson, M. R., 225

 Taguchi, G., 463-464, 477
 Tamhane, A. C., 225-226, 228, 252, 333, 335
 Tang, P. C., 184
 Throckmorton, T. N., 109
 Tobias, R. D., 225
 Tukey, J. W., 200, 224, 226, 302-303, 305, 387-388, 440

 van Belle, G., 20, 313, 319
 van Eijnsbergen, A. C., 306
 Venn, J., 25
 Vining, G. G., 498, 512

 Voss, D. T., 134

 Wagenaar, W. A., 402, 547
 Walters, D. E., 575
 Wampler, J. L., 497
 Wang, Y. C., 387
 Ward, G. C., 303
 Watson, G. S., 127
 Webb, D. F., 511
 Welch, B. L., 177, 286, 384-385
 Welsch, R. E., 258
 Westfall, P. H., 225
 White, R. F., 109
 Wilk, M. B., 161, 300, 317, 323-324, 386-387
 Williams, E. J., 401, 403, 547
 Wilson, K. B., 497, 499, 506, 518-519
 Wolfinger, R. D., 225

 Yates, F., 59, 221, 290-291, 295, 330, 377, 394-396, 420, 468, 534, 549
 Yeh, C. M., 341-342
 Youden, W. J., 394
 Young, S., 479

 Zahn, D. A., 441
 Zedek, S., 548
 Zyskind, G., 107-109, 127, 323

Subject Index

- Additivity
 - in the broad sense, 161, 164, 218, 286, 300, 329, 385, 425
 - in the strict sense, 158, 161, 164, 171, 218, 280, 286, 300, 377
 - transformation to, 388
 - unit-treatment, 40, 197, 329, 377, 425, 537
- Aitken
 - equation, 125-127
 - theorem, 125-127
- Alias
 - matrix, 516
 - structure, 454-455, 473-475
- Analysis
 - approximate, 295
 - cluster, 229
 - of data, 29
 - of experiments, 20-22, 26
 - mixed model, 578
 - nonparametric, 68
 - regression, 220, 497. *See also*
 - Regression, analysis
 - statistical, 30, 36, 43, 55-57
 - trend, 575
- Analysis of covariance, 118, 121, 239, 242, 258-260, 292, 295, 389
 - algebra of, 121
 - model, 118
 - with subsampling, 258
 - table, 248-249, 294
 - technique, 295
- Analysis of variance (ANOVA), 36, 42-46, 68, 79, 92, 95, 99-101, 112, 115, 130, 151, 165-166, 177
- auxiliary, 122, 242
- BIBD, 361-364
- between-and-within subjects design, 566
- CRD, 167
 - with orthogonal contrasts, 217, 233
 - with orthogonal polynomials, 224, 235
 - with repeated measures, 581
 - with subsampling, 194
 - with unequal numbers, 182
- cross-over design, 410
- first-order design, 501-502, 525
- GRBD, 317-318, 324, 357
- Græco-Latin square, 397
- IBD, 330
- LSD, 383, 405
 - replicated, 391-393, 408
- orthogonal, 93-94
- preliminary, 433
- RCBD, 282-284, 344
 - with crossed blocking factors, 311
 - with nested blocking factors, 309-310, 354
 - under nonadditivity, 304
 - with subsampling, 289, 310, 349, 354
- second order design, 505
- SPD(BIBD, RCBD), 555
- SPD(CRD, BIBD), 551

- SPD(CRD, LSD), 548
- SPD(CRD, RCBD), 545
 - with repeated measures, 589
- SPD(GRBD, BIBD), 553
- SPD(GRBD, IBD), 554
- SPD(GRBD, RCBD), 552
- SPD(LSD, RCBD), 549
- SPD(RCBD, GRBD), 556
- SPD(RCBD, RCBD), 536
- split-block design, 558
- split-plot design, 536, 563
- split-split-plot design, 561
- table, 245-246, 259, 301, 430-431, 438, 448, 467
- two-way classification, 105
- uniformity trial, 291, 544
- Anaximander, 12
- Anaximenes, 12
- Aphorism, 3, 7,
 - existential, 9
- Aquinas, 13, 17
- Aristotle, 9, 12-13, 17
- Arrangement(s)
 - random, 340
 - systematic, 341
- Associates
 - first, 336
 - second, 336
- Association scheme, 336
- Avicenna, 13
- Bacon, 4-7, 13, 19
- Balance, 102
- Balancedness, 59
- Basis
 - inductive, 420
 - inferential, 64
- Bayes
 - approach, 123
 - empirical, 139
 - hierarchical, 123
- Belief(s), 141, 144
 - calculus, 26
 - rational, 25
- Bernoulli, 25
 - random variable, 155
 - trial, 25
- Bertrand, 25
- Bias, 514-515
 - squared, 515
- Block(s)
 - design, *see* Design
 - effect, *see* Effect, block
 - incomplete, 328, 446
 - size, 329
- Blocking, 61
 - effectiveness of, 288
 - factors, *see* Factor(s)
 - orthogonal, 390, 507-508, 511
 - in two directions, 390
- Bonferroni
 - inequality, 225
 - procedure, 225, 228
 - t*-statistic, 225-226
- Borel, 25
- Boyle, 13
- Causality, 16-19
- Causation, *see* Causality
- Cause, *see* Causality
- Central limit theorem, 148
- Classification
 - Aristotelian, 5
 - models, 74-75
 - one-way, 102-103, 109
 - two-fold nested, 109, 112, 193
 - two-way, 95, 103-106, 109, 112-116, 329
- Collinearity, 263
- Comparison(s), *see also* Contrast(s)
 - a priori, 213
 - with a control, 227
 - multiple, 224, 251-252, 269, 286
 - orthogonal, 214-215, 218
 - pairwise, 224
 - post-hoc, 225
 - preplanned, 213, 219, 250
 - split-plot treatment, 543-544
 - treatment, 59, 165, 250, 332, 377
 - estimated, 218
 - whole-plot treatment, 543-544, 549

- Component(s)
 covariance, 133, 578
 design, 56
 error, 56
 treatment, 56
 variance, 133, 578
- Comprehensiveness, 420
- Computer packages, *see* SAS
- Confidence interval(s), 27, 542
 estimation, 37
 simultaneous, 226-228
- Confounding
 complete, 449
 partial, 449, 452
 system of, 447, 457, 472, 475, 507-508, 511, 552-554
- Connectedness, 59, 329
- Contrast(s), *see also* Comparison(s)
 among treatment effects, 160
 cubic, 223
 defining, 455-456
 orthogonal, 214-215, 424
 complete set of, 214
 quadratic, 223
 single-d.f., 441-442
 standardized, 215
 treatment, 214, 282, 380
- Control
 historical, 139, 148
 local, 45, 278
 statistical, 24
- Coordinate
 paper, 520
 system, 519
- Copernicus, 13
- Correlation, epistemic, 11-12
- Covariate(s), 76, 240, 292
 affected by treatments, 256
 multiple, 259, 263
- Darwin, 3, 9
- Data
 analysis, 5, 11-12, 15, 100, 122
 collection, 34, 38
 longitudinal, 573
 observational, 137
 snooping, 227
 structure, 100, *See also* Structure, data
- Deduction, 6-7
- Degrees of freedom (d.f.), 79, 166, 183
 denominator, 195
 loss of, 290
- Democritus, 12
- Descartes, 9, 12
- Design, *see also* Experiment(s),
 all-bias, 517
 all-variance, 517
 axial, 506-508, 520-521
 balanced incomplete block (BIBD),
 63, 330-335, 359, 304, 549, 552-554
 balanced residual effect, 401
 balanced treatment incomplete block (BTIBD), 63, 333
 between-and-within subjects, 545-548
 binary, 330, 342
 Box-Behnken, 65, 506
 carry-over, 397
 central composite (CCD), 506-508, 512-514, 524, 527-529
 change-over, 397, 400
 complete block, 341
 completely randomized (CRD),
 34, 62, 106, 151-153, 218, 529
 with subsampling, 191
 with unequal numbers, 179, 219
 connected, 339
 construction of, 23
 counterbalanced, 397, 545
 cross-over, 63, 397, 573
 diagram-balanced, 545
 disconnected, 339
 economical, 511
 effects, *see* Effects,
 equireplicate, 330, 343
 equivalent estimation, 512
 error-control, 33-41, 45, 53-56, 61-65, 153, 277-278, 298, 328,

- 377, 390, 421, 475, 497-499, 573, 580
 - error-reduction, 277, 419, 446
 - of experiments, 1, 16, 20-22, 26, 29
 - extended block, 63, 337-338, 554
 - factorial, *see* Factorial(s)
 - first-order, 506-507, 513, 525
 - generalized randomized block (GRBD), 63, 106, 312-314, 323, 353, 554
 - Græco-Latin square, 63-64, 395-396
 - incomplete block (IBD), 34, 65, 113
 - 134, 295, 328-329, 421-422, 446, 457, 509, 549, 552-554
 - Kronecker product, 477
 - Latin rectangle, 62, 393-394
 - Latin square (LSD), 34, 45, 62-64, 149-151, 374, 377, 387. *See also* Latin square(s)
 - extended, 395
 - incomplete, 63-64, 394-395
 - mutually orthogonal, 63, 395-396
 - replicated, 63, 390, 548
 - type, 62-63, 373-375
 - lattice, 63-65
 - mixture, 519
 - model, 260
 - moments, 506, 515-516
 - nonorthogonal, 332, 422
 - observation, 333-334, 573
 - optimal, 517
 - orthogonal, 421, 425, 504-506, 539
 - pairwise balanced, 338
 - parameter, 477
 - partially balanced incomplete block (PBIBD), 63, 335-337, 552
 - proper, 330, 342
 - randomized block, 34, 62-63, 277
 - replicated, 62-63
 - randomized complete block (RCBD), 34, 63, 151, 278, 428, 543
 - with subsampling, 288
 - repeated measures, 65, 397, 545, 573, 577
 - balanced, 401
 - resolution III, 455, 475
 - resolution IV, 457, 475
 - resolution V, 457, 475, 505
 - resolvable, 511
 - response surface, 65, 497. *See also* Response surface
 - row-column, 34, 374
 - sampling, 33-37, 61
 - second-order, 504-505, 513
 - simplex, 504
 - centroid, 520-521
 - lattice, 520
 - split-block, 555
 - incomplete, 560
 - split-plot, 38, 45, 512, 528, 534, 539
 - in strips, 555
 - in time, 545, 577
 - type, 39, 65, 151, 511, 533
 - split-split-plot, 511, 560
 - statistical, 30
 - switch-over, 397
 - systematic block, 340
 - treatment, 33-37, 45, 61, 64-65, 390, 419, 499, 573
 - trend-free block, 341-342
 - nearly, 342
 - unbalanced, 511
 - unbiased, 382, 387
 - variance balanced, 59, 338
 - Williams, 402
 - Youden square, 34, 394-395
- Dewey, 19
- Diagrammatic representation, 109
- Difference(s)
- minimum, 184
 - standardized, 184
 - smallest detectable, 184
 - treatment, 184, 242
 - adjusted, 242
- Distribution
- beta, 175-177
 - chi-squared, 129, 303
 - noncentral, 131

- F -, 132, 177
 - central, 132
 - noncentral, 132, 183
- joint, 26, 72, 130
- mathematical, 10
- noninformative prior, 26
- normal (Gaussian), 10, 72
 - multivariate, 129-130
- posterior, 25-26
- prior, 26
- probability, 26, 139
- randomization, 177. *See also*
 - Randomization
- studentized range, 226
- of sum of squares, 130
- t -, 131, 542
 - noncentral, 131
- theory, 129
- Effect(s)
 - block, 37-38, 134
 - random, 324
 - carry-over, 398, 403
 - design, 37, 40, 134
 - direct, 398-400
 - error, 37, 40, 57
 - fixed, 325
 - interaction, 37-38, 419, 468
 - learning, 545
 - linear, 505
 - linear \times linear, 505
 - main, 37-38, 419-425, 468-470
 - multiplicity, 224
 - order, 548
 - quadratic, 505
 - random, 57, 134
 - residual, 398-400
 - second-order, 403
 - second order, 513, 517
 - simple, 423
 - size, 184
 - systematic, 57
 - treatment, 37, 40, 46-47
 - differential, 171-173
 - estimated (ERE), 290-292, 389, 543
- Efficiency, 59
 - factor, 333, 336
 - relative (RE), 288-291, 389, 543
- Einstein, 13
- Equation(s)
 - Aitken, 125-127
 - like, 133, 578
 - linear, 81
 - normal (NE), 77, 80-83, 87, 115-117, 125 242, 246, 254, 293, 329, 500
 - conjugate, 77, 129
 - reduced (RNE), 90, 121, 329, 360
 - theory of, 81
- Error
 - components, 162
 - estimation of, 137, 148
 - experimental, 37-48, 68, 163-165, 191, 258
 - mean squared, 514
 - integrated (IMSE), 516
 - measurement, 10, 14, 23, 39-40, 161-162
 - observational, 24, 39-48, 68, 191, 315
 - pure, 430, 502, 505
 - rates, 224
 - comparisonwise, 224-225
 - familywise, 224-225
 - sampling, 39, 42, 48, 68, 161-162
 - selection, 161-162
 - space, 129
 - split-plot, 512
 - standard (se), 151
 - state, 161-162
 - structure, 162
 - technical, 161, 164
 - treatment, 38, 161-162, 315
 - unit, 40, 160, 239
 - variance, 515
 - whole-plot, 512
- Estimability, 77, 123
- Estimate
 - ANOVA-type, 133
 - best linear unbiased (BLUE), 125-27, 160, 242

- of error, 137, 148
 - generalized least squares (GLS), 126
 - interval, 165
 - M*-, 228, 257
 - ordinary least squares (OLS), 126-127, 160
 - point, 165
- Estimation, *see also* Estimate
 - space, 129
- Estimator, *see* Estimate
- Expectation, posterior, 25
- Experiment(s), *see also* Experimentation, Studies
 - absolute, 22, 65, 73
 - agricultural, 148
 - agronomic, 65, 420, 549, 556
 - arithmetical, 142
 - comparative, 14-15, 19-24, 65, 71-73, 151, 497, 523
 - confirmatory, 29-30
 - design of, 16, 20-22, 26, 29
 - exploratory, 29-30, 475
 - factorial, 59, 64, 419-422, 533
 - asymmetrical, 64
 - fractional, *see* Factorial(s)
 - symmetrical, 64
 - investigative, 144
 - Lady tasting tea, 139-140
 - mixture, 519, 523
 - noisy, 142
 - psychological, 545
 - randomized, 145
 - replicated randomized block, 307
 - triangular, 140
 - types of, 23
- Experimentation
 - industrial, 30, 420, 497, 511, 543
 - scientific, 30, 46
 - sequential, 43
- Explanation, 18
- Factor(s)
 - between-subjects, 545
 - blocking, 32, 35, 106, 278, 306, 313, 373, 440, 580
 - crossed, 308
 - nested, 308
 - classification, 32
 - confounded, 101
 - correction, 94
 - crossed, 101
 - easy-to-change, 511-512, 543
 - efficiency, 333, 336
 - hard-to-change, 511-513, 524, 543
 - intrinsic, 35, 38-42, 45, 51-53, 56, 106, 134, 278, 313-314, 325, 373, 440, 552
 - level, 54
 - nested, 106
 - nonspecific, 35, 38-42, 45, 56, 134, 278, 373
 - qualitative, 52
 - quantitative, 52
 - split-plot, 534
 - treatment, 32-35, 42, 51-53, 422
 - whole-plot, 534
 - within-subjects, 545
- Factorial(s), *see also* Design, Experiment(s)
 - asymmetrical, 66, 476, 479
 - complete, 511
 - fractional, 64-66, 453-455, 462-, 463, 472, 475, 479, 505-506
 - of resolution III, 503
 - of resolution IV, 503
 - of resolution V, 503, 511
 - full, 503
 - highly fractionated, 475
 - mixed, 476, 548
 - pure, 476
 - symmetrical, 66, 476
 - 2^n , 422, 446, 462, 503, 509
 - 3^n , 465, 472, 505, 509
- Faraday, 13
- Fit
 - badness of, 76
 - lack of, 223, 502, 505, 524
- Frequency(ies)
 - proportional, 96
 - relative, 10
- Frequentist approach, 123

- Function(s)
 - estimable, 78, 81, 125, 131, 242, 459, 557
 - identifiable, 81
 - likelihood, 26
 - linear, 130
 - parametric, 137
 - polynomial, 498
 - quadratic, 130, 166
- Games of chance, 25
- Gauss-Markov
 - linear model (GMLM), 124-125
 - normal linear model (GMNLM), 128-131, 137, 147
 - properties, 147
 - theorem, 124
- Half-normal plot, 441-443
- Heisenberg uncertainty principle, 2, 11
- Heraclitus, 12
- Heterogeneity, 239
 - elimination of, 395
 - of experimental units, 160
- Homogeneity, 193
 - of groups, 229
- Hume, 9
- Huynh-Feldt condition, 545, 577
- Hypothesis, 6-7
 - falsification of, 7
 - reductionist, 14
 - research, 32, 41-44, 53
 - statistical, 32, 43
 - working, 32
- Identifiability, 76, 114, 120
- Induction, 4, 6-7
- Inequality
 - Bonferroni, 225
 - Tchebycheff, 139
- Inference, 16
 - Bayesian, 26
 - statistical, 24, 57, 122, 151
 - types of, 7, 36
- Information
 - inter-block, 134, 553
 - loss of, 478
 - supplementary, 59, 239-242, 248, 292
- Interaction(s), 419, 470, 475
 - antidirectional, 319
 - antagonistic, 319
 - block-treatment, 278, 300-302, 306-308, 312-314, 317-319, 338
 - codirectional, 313, 319
 - components, 468-470, 475
 - effects, *see* Effects, interaction
 - first-order, 421
 - generalized, 449, 456, 474-475
 - higher order, 420-421, 428
 - linear \times linear, 505
 - lower order, 420
 - plot, 320-321
 - replication \times treatment, 391-392
 - row-column, 379
 - simple, 424
 - synergistic, 313
 - three-factor, 457, 471, 504
 - treatment \times design, 134
 - treatment-time, 575, 580
 - two-factor, 421, 456, 471, 504
 - unit-treatment, 300-301, 314
- Interval(s)
 - confidence, 27, 217
 - estimation, 37
 - simultaneous, 226-227
 - statistical, 137
- Intervention, *see* Studies, intervention
- Jeffreys, 26
- Kant, 9, 17
- Keeton, 17
- Kepler, 4, 13
- Knut Vik square, 149-150
- Lagrange multipliers, 87
- Latin rectangle, 393-394
- Latin square(s), 62, 376, 548
 - completely counterbalanced, 402
 - complete orthogonalized, 397
 - cyclic, 402

- design (LSD), *see* Design, Latin square
- diagram-balanced, 547
- Græco-, 396
- incomplete, 394, 402
- mutually orthogonal (MOLS), 397, 403
- orthogonal, 395
- principle, 62-64, 390, 393
- reduced, 376
- Lavoisier, 13
- Law(s)
 - Kepler's, 4
 - Mendel's, 4
 - of succession, 25
- Least squares
 - analysis, 293-297, 335, 500
 - fitting, 76, 80, 86, 119
 - generalized (GLS), 126, 512
 - mean (LSM), 244, 325-326, 332, 422
 - method of, 37, 57, 77, 220-221, 242, 466
 - ordinary (OLS), 126, 512
- Leucippus, 132
- Level(s)
 - coded, 500
 - equidistant, 219
 - significance, 172-173, 177-179, 183
- Likelihood, 11
 - function, 26
 - residual maximum (REML), 580
- Linear model, 37-38, 44-46, 71-73
 - affine, 85-86
 - approximative, 77
 - classificatory, 74-75
 - conditional, 85, 99
 - derived, 68, 127, 159, 164, 278, 314, 537
 - functional, 74
 - Gaussian, 26
 - Gauss-Markov (GMLN), 124-125, 128-131, 137, 147
 - k-part, 97
 - ordered, 90-94
 - stochastic, 74, 77, 123
 - theory, 71
 - 2-part, 90
 - 3-part, 94, 329
- Locke, 9
- Logic, Aristotelian, 9
- Loss
 - of degrees of freedom, 290
 - of information, 253
 - of power, 290
 - of sensitivity, 290
- Mathematics, foundations of, 6
- Matrix
 - design, 506
 - design-model, 466, 504
 - generalized inverse, 81-83, 125, 332
 - idempotent symmetric (s.i.p), 78, 84, 87
 - incidence, 118, 329, 333, 339, 509
 - information, 330
 - model, 73
 - Moore-Penrose (M-P) inverse, 84-86, 124
 - orthogonal, 126, 216
 - projection, 91
 - variance-covariance, 124-125, 578
- Maxwell, 13
- Mean, admissible, 107-108, 115
- Mean square(s)
 - expected, 168
 - synthetic error, 326-327
- Measure(s),
 - repeated, 23, 573-574, 578
 - summary, 575
- Measurement(s)
 - process, 10, 22-24
 - repeated, 23-24, 573
 - scale of, 34, 197
 - variability, 24
- Mendel, 4
- Method(s)
 - ANOVA, 580
 - delta, 198
 - of parallel tangents (PARTAN), 519
 - of statistical differentials, 198
 - of steepest ascent, 518

- Mill, 18
 Model(s), *see also* Linear model
 approximate, 71, 74
 classificatory, 34, 74
 conditioned, 85-87
 first-order, 500
 fitting a, 76
 fixed, 132-133, 323
 full, 430
 means, 99, 114
 misspecification, 513
 mixed, 132-134, 325
 multiplicative, 303
 nonlinear, 34
 nonorthogonal, 332
 overparameterized, 100, 115
 partitioned, 94
 polynomial, 519
 probability, 10
 random, 132-133, 325
 randomization, 159
 regression, 34
 first-order, 500
 second-order, 504
 relative frequency, 10
 statistical, 30, 34
 testing of, 7
 stochastic, 74, 128, 138
 subject matter, 32, 35, 51, 54
 subsampling, 191
 three-part, 94
 two-way classification, 329
 well-formulated, 109-110, 115
 Monte Carlo studies, 178
 Multicollinearity, 74-75

 Newton, 17
 Nonadditivity, 196, 300-302, 312, 386-387
 testing for, 303
 Noncentrality parameter, 131-132, 183, 195
 Nonorthogonality, 400
 Normality assumption, 257

 Observation(s)
 adjusted, 241
 high-leverage, 258
 missing, 55, 295-298, 389
 estimated, 297
 multivariate, 34
 process, 1-2, 10
 supplementary, 258
 types of, 3
 univariate, 34
 validation of, 2
 Optimality, 59
 A-, 59
 D-, 59
 E-, 59
 Orthogonal array, 463-464
 Orthogonality, 59, 218, 400-402

 Period,
 extra, 400
 pre-, 400
 wash-out, 398
 Plan
 main effect, 455, 480
 orthogonal, 462-463
 saturated, 464
 Plato, 12-13
 Plot(s)
 half-normal, 441-443
 interaction, 320-321
 split, 534, 537, 548, 556
 split-split, 560
 whole, 534, 537, 548, 556
 Poincaré, 8, 25
 Points,
 axial, 506-508
 center, 506-508
 factorial, 506-508
 Polynomial(s)
 canonical, 521-522
 first-order, 499-500
 low-order, 499
 orthogonal, 220-222, 342, 441, 505, 576
 Tchebycheff, 221
 Popper, 7-8
 Population,

- marginal mean, 245
 - reference, 323
 - target, 32, 60
- Power
 - explanatory 15
 - of F -test, 182
 - loss of, 290
 - transformation, 200
- Precision
 - increase in, 253
 - of treatment comparisons, 278
- Predictive margin, 245
- Principle(s)
 - of blocking, 34
 - of experimentation, 29
 - of indifference, 25
 - Latin square, 62-64, 390, 393
 - split-unit, 533
- Probability, 141
 - conditional, 25
 - continuous, 10
 - degrees of, 25
 - frequency theory of, 25-26
 - joint, 24
 - structure, 25
 - theory, 72
- Procedure(s)
 - Bonferroni, 225, 228
 - Calinski-Corsten, 231
 - Dunnett's, 228
 - hierarchical agglomerative, 229
 - hypothesis falsification, 7
 - Johnson-Graybill, 306
 - Mandel's, 302
 - multiple comparison, 224, 250-251
 - nonparametric, 228
 - optimization, 518
 - Satterthwaite, 326, 561
 - Scheffé, 227-228
 - stepwise, 229
 - studentized range, 226, 229
 - Tukey, 226, 252
 - Tukey-Kramer, 226, 252, 269
- Process
 - control, 30
 - evolutionary, 1
 - manufacturing, 24
 - measurement, 10, 34
 - observational, 1-2, 10
 - production, 30
 - randomization, *see* Randomization, process
 - of science, 1
 - sequential, 30
 - stochastic, 26, 72
- Projection(s)
 - matrix, 91
 - orthogonal, 91
- Projector, orthogonal, 91
- Protocol
 - experimental, 55, 139
 - measurement, 10
 - observation, 2
- Pythagoras, 12
- Quadratic form, 128
- Quality control, off-line, 477
- Ramsey, 26
- Randomization, 26, 34, 45, 55-56, 61, 106, 137, 140-141, 147-151, 278, 376
 - analysis, 68-69, 180
 - distribution, 26
 - independent, 534, 538
 - procedures, 27, 68, 111, 156-157, 180, 377, 380, 533, 548
 - process, 154, 157, 164, 171, 280-, 281, 315, 533, 537
 - repetitions of, 158
 - restricted, 280, 291
 - test, 26, 69, 150, 172-173, 180, 285, 385
 - approximation to the, 69, 173-174, 193, 217, 248, 286-288, 317, 538
 - theory, 134, 177, 287, 303, 382
 - unrestricted, 291
- Random numbers, 154
- Random variable(s), 22-24
 - Bernoulli, 155

- design, 68, 154-155, 158, 280, 379, 537
- Gaussian, 22
- normal, 129
 - multivariate, 129
- Range of validity, 60
- Region
 - experimental (ER), 498, 503
 - operational (OR), 498
- Regression
 - analysis, 220, 497, 502
 - coefficient, 221, 304, 466-467, 505, 528
 - line, 302
 - polynomial, 262
 - second-order linear, 504
- Relation, *see* Relationship
- Relationship
 - defining, 455-456
 - functional, 498
 - identity, 455, 475, 479
- Reparameterization, 81
- Repetition(s), 139
 - population of, 137-138
- Replication(s), 45, 61
 - fractional, 451-453
 - number of 180, 184, 186-190, 193-195
 - effective, 451
 - unequal, 179-180
- Residual, *see* Error
- Response(s)
 - conceptual, 157-158, 161, 281, 315, 379, 537
 - curve, 497, 514
 - observed, 315, 379
 - optimum, 499, 503
 - predicted, 501
- Response surface, 497-500
 - design, 497-499
 - first-order, 503
 - methodology (RSM), 497-499, 519, 523
 - second-order, 509
- Rightmost bracket, 107, 111
- Rotatability, 506
- Sartre, 9
- SAS, 69
 - PROC FACTEX, 451, 459, 486-491
 - PROC GLM, 201, 230-232, 264, 269, 343, 348, 353, 404, 407, 430, 443, 446, 481, 523, 562, 580
 - PROC IML, 86
 - PROC MIXED, 201, 343, 348, 353, 483, 523, 562, 568, 578-580
 - PROC PLAN, 154, 180, 278-279, 315-316, 377-378, 481
 - PROC POWER, 185
 - PROC REG, 523
 - PROC RSREG, 523-524
- Science(s)
 - descriptive, 9
 - exact, 12
 - general, 14
 - history, 5
 - physical, 14
 - process of, 1
 - type of, 9
- Scientific objective, 58
- Scope of validity, 277
- Sensitivity of experiment, 45
- Simplex
 - coordinate system, 519
 - design, 504
 - k*-dimensional, 504, 519
- Slope(s)
 - common, 253
 - equality of, 269
- Socrates, 6, 12
- Space
 - column, 91
 - error, 129
 - estimation, 129
 - row, 76
- Statistical Analysis System, *see* SAS
- Statistical software, *see* SAS
- Statistical triangle, 46
- Structure(s)
 - alias, 454-455, 473-475

- blocking, 45
 - nested, 353, 428, 440
- classificatory, 100, 106
 - balanced, 100-101
- correlation, 577
 - estimated, 588
- covariance, 127, 160, 164, 168, 545, 575, 578
 - compound symmetry, 577-580
 - first-order autoregressive, 578-580
 - spatial power, 579
 - unstructured, 579-580
- data, 100
 - balanced, 107-108, 111-112,
 - classificatory, 99, 118
 - unbalanced, 112
- diagram(s), 110-111
- error, 56
- factor balanced, 112
- factorial, 42, 64, 106, 419-421, 440, 543, 552. *See also* Factorial(s)
 - asymmetrical, 64
 - symmetrical, 64
- Latin square, 380
- variance-covariance, 127
- Studies, *see also* Experiment(s)
 - experimental, 104-106, 138, 149,
 - intervention, 106, 137-138, 149
 - observational, 104-106, 134, 137-138, 149
 - preliminary, 185, 191
 - simulation, 177, 286. *See also* Monte Carlo studies
- Subject matter knowledge, 328, 422
- Subsample, size of, 195
- Subsampling, 34, 40, 67-68, 191-193, 288, 353
- Sub-subsampling, 67-68
- Sum(s) of squares, 37, 426
 - partial, 98, 550
 - sequential, 98
 - Type I, 98, 330
 - Type III, 98
- Syllogism, 5-6
 - basic, 5-6
- Symmetry, compound, 577-580
- Synergism, 43
- Taylor series expansion, 421
- Test(s)
 - Bonferroni, 225
 - criterion, 151, 172
 - Duncans multiple range, 226, 251
 - F -, 151, 174, 177, 217, 285, 502, 538
 - power of, 182
 - Fishers protected LSD, 225
 - F -max, 323
 - of hypotheses, 7, 37, 57, 131, 171
 - lack-of-fit, 223, 502-505
 - preliminary, 313, 323
 - randomization, 69, 177, 285. *See also* Randomization, test
 - randomized triangular, 140
 - significance, 7, 37, 148-151, 165, 542
 - size of, 183
 - statistical, 137
 - studentized range, 226
 - t -like, 257
 - treatment, 226-227
 - Tukey's, 303-305, 388
- Tetrahedron, 504
- Thales of Miletus, 12
- Theorem
 - Aitken, 125-127
 - central limit, 148
 - Gauss-Markov, 125-126
- Theory
 - axiom, 13
 - development of, 4, 10
 - falsifiable, 14
 - Gauss-Markov normal linear model (GMNLM), 174-176
 - mathematical, 11
 - normal, 285-286

- randomization, 176-177, 285-286. Tycho Brahe, 13
 - See also* Randomization,
 - theory
 - scientific, 8
 - statistical, 11
 - types of, 11
- Time series, 17
- Transformation(s), 196-199, 312
 - to additivity, 388
 - power, 200
 - sets, 376-377, 384
- Treatment(s)
 - combinations, 420
 - control, 227
 - design, *see* Design, treatment
 - factorial, 64
 - mean,
 - adjusted, 244
 - qualitative, 52, 213, 497
 - quantitative, 52, 219, 497
 - split-plot, 539-540, 543, 560
 - effect, 539
 - split-split-plot, 560, 561
 - test, 227
 - whole-plot, 539-540, 543, 560
 - effect, 539-540
- Trend, 340
 - analysis, 575
 - linear, 223, 341-343
 - overall, 577
- Trial
 - agronomic, 278
 - Bernoulli, 25
 - binomial, 24
 - randomized clinical, 32, 35
 - uniformity, 290-291, 543
- Triangle, equilateral, 504
- Unbiasedness, 59
- Unit(s)
 - error, 160
 - experimental (EU), 20, 34, 38, 68, 138, 153, 533
 - observational (OU), 34, 38, 68
 - sampling, 68
- Variability, *see* Variation
- Variable(s)
 - classificatory, 118
 - coded, 514
 - concomitant, 15, 76, 118
 - explanatory, 35-38, 74, 151
 - function of, 71
 - mathematical, 11-12, 22
 - process, 523
 - random, *see* Random variable(s)
 - regressor, 302
 - response, 35-36, 52
- Variance(s), 514
 - average, 251-252, 336
 - estimator, 248, 562
 - experimental error, 317, 557
 - component, 163, 193, 288
 - nonconstancy of, 197
 - observational error, 317
 - component, 163, 193, 288
 - prediction, 506
- Variation, 26
 - induced, 277
 - random, 38, 239
 - sources of, *see* Analysis of variance, table
 - systematic, 45, 62, 239

This Page Intentionally Left Blank

WILEY SERIES IN PROBABILITY AND STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J. B. Kadane, Geert Molenberghs, David W. Scott, Adrian F. M. Smith, Sanford Weisberg*

Editors Emeriti: *Vic Barnett, J. Stuart Hunter, David G. Kendall, Jozef L. Teugels*

The *Wiley Series in Probability and Statistics* is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches.

This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

- † ABRAHAM and LEDOLTER · Statistical Methods for Forecasting
- AGRESTI · Analysis of Ordinal Categorical Data
- AGRESTI · An Introduction to Categorical Data Analysis, *Second Edition*
- AGRESTI · Categorical Data Analysis, *Second Edition*
- ALTMAN, GILL, and McDONALD · Numerical Issues in Statistical Computing for the Social Scientist
- AMARATUNGA and CABRERA · Exploration and Analysis of DNA Microarray and Protein Array Data
- ANDÉL · Mathematics of Chance
- ANDERSON · An Introduction to Multivariate Statistical Analysis, *Third Edition*
- * ANDERSON · The Statistical Analysis of Time Series
- ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG · Statistical Methods for Comparative Studies
- ANDERSON and LOYNES · The Teaching of Practical Statistics
- ARMITAGE and DAVID (editors) · Advances in Biometry
- ARNOLD, BALAKRISHNAN, and NAGARAJA · Records
- * ARTHANARI and DODGE · Mathematical Programming in Statistics
- * BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
- BALAKRISHNAN and KOUTRAS · Runs and Scans with Applications
- BALAKRISHNAN and NG · Precedence-Type Tests and Applications
- BARNETT · Comparative Statistical Inference, *Third Edition*
- BARNETT · Environmental Statistics
- BARNETT and LEWIS · Outliers in Statistical Data, *Third Edition*
- BARTOSZYNSKI and NIEWIADOMSKA-BUGAJ · Probability and Statistical Inference
- BASILEVSKY · Statistical Factor Analysis and Related Methods: Theory and Applications
- BASU and RIGDON · Statistical Methods for the Reliability of Repairable Systems
- BATES and WATTS · Nonlinear Regression Analysis and Its Applications

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- BECHHOFFER, SANTNER, and GOLDSMAN · Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons
- BELSLEY · Conditioning Diagnostics: Collinearity and Weak Data in Regression
- † BELSLEY, KUH, and WELSCH · Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
- BENDAT and PERSOL · Random Data: Analysis and Measurement Procedures, *Third Edition*
- BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
- BERNARDO and SMITH · Bayesian Theory
- BHAT and MILLER · Elements of Applied Stochastic Processes, *Third Edition*
- BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
- BILLINGSLEY · Convergence of Probability Measures, *Second Edition*
- BILLINGSLEY · Probability and Measure, *Third Edition*
- BIRKES and DODGE · Alternative Methods of Regression
- BISWAS, DATTA, FINE, and SEGAL · Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics
- BLISCHKE AND MURTHY (editors) · Case Studies in Reliability and Maintenance
- BLISCHKE AND MURTHY · Reliability: Modeling, Prediction, and Optimization
- BLOOMFIELD · Fourier Analysis of Time Series: An Introduction, *Second Edition*
- BOLLEN · Structural Equations with Latent Variables
- BOLLEN and CURRAN · Latent Curve Models: A Structural Equation Perspective
- BOROVKOV · Ergodicity and Stability of Stochastic Processes
- BOULEAU · Numerical Methods for Stochastic Processes
- BOX · Bayesian Inference in Statistical Analysis
- BOX · R. A. Fisher, the Life of a Scientist
- BOX and DRAPER · Response Surfaces, Mixtures, and Ridge Analyses, *Second Edition*
- * BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
- BOX and FRIENDS · Improving Almost Anything, *Revised Edition*
- BOX, HUNTER, and HUNTER · Statistics for Experimenters: Design, Innovation, and Discovery, *Second Edition*
- BOX and LUCENO · Statistical Control by Monitoring and Feedback Adjustment
- BRANDIMARTE · Numerical Methods in Finance: A MATLAB-Based Introduction
- † BROWN and HOLLANDER · Statistics: A Biomedical Introduction
- BRUNNER, DOMHOF, and LANGER · Nonparametric Analysis of Longitudinal Data in Factorial Experiments
- BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
- CAIROLI and DALANG · Sequential Stochastic Optimization
- CASTILLO, HADI, BALAKRISHNAN, and SARABIA · Extreme Value and Related Models with Applications in Engineering and Science
- CHAN · Time Series: Applications to Finance
- CHARALAMBIDES · Combinatorial Methods in Discrete Distributions
- CHATTERJEE and HADI · Regression Analysis by Example, *Fourth Edition*
- CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
- CHERNICK · Bootstrap Methods: A Guide for Practitioners and Researchers, *Second Edition*
- CHERNICK and FRIIS · Introductory Biostatistics for the Health Sciences
- CHILÈS and DELFINER · Geostatistics: Modeling Spatial Uncertainty
- CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies, *Second Edition*
- CLARKE and DISNEY · Probability and Random Processes: A First Course with Applications, *Second Edition*
- * COCHRAN and COX · Experimental Designs, *Second Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- CONGDON · Applied Bayesian Modelling
 CONGDON · Bayesian Models for Categorical Data
 CONGDON · Bayesian Statistical Modelling
 CONOVER · Practical Nonparametric Statistics, *Third Edition*
 COOK · Regression Graphics
 COOK and WEISBERG · Applied Regression Including Computing and Graphics
 COOK and WEISBERG · An Introduction to Regression Graphics
 CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Third Edition*
 COVER and THOMAS · Elements of Information Theory
 COX · A Handbook of Introductory Statistical Methods
 * COX · Planning of Experiments
 CRESSIE · Statistics for Spatial Data, *Revised Edition*
 CSÖRGÖ and HORVÁTH · Limit Theorems in Change Point Analysis
 DANIEL · Applications of Statistics to Industrial Experimentation
 DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Eighth Edition*
 * DANIEL · Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*
 DASU and JOHNSON · Exploratory Data Mining and Data Cleaning
 DAVID and NAGARAJA · Order Statistics, *Third Edition*
 * DEGROOT, FIENBERG, and KADANE · Statistics and the Law
 DEL CASTILLO · Statistical Process Adjustment for Quality Control
 DEMARIS · Regression with Social Data: Modeling Continuous and Limited Response Variables
 DEMIDENKO · Mixed Models: Theory and Applications
 DENISON, HOLMES, MALLICK and SMITH · Bayesian Methods for Nonlinear Classification and Regression
 DETTE and STUDDEN · The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis
 DEY and MUKERJEE · Fractional Factorial Plans
 DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
 DODGE · Alternative Methods of Regression
 * DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
 * DOOB · Stochastic Processes
 DOWDY, WEARDEN, and CHILKO · Statistics for Research, *Third Edition*
 DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
 DRYDEN and MARDIA · Statistical Shape Analysis
 DUDEWICZ and MISHRA · Modern Mathematical Statistics
 DUNN and CLARK · Basic Statistics: A Primer for the Biomedical Sciences, *Third Edition*
 DUPUIS and ELLIS · A Weak Convergence Approach to the Theory of Large Deviations
 EDLER and KITSOS · Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment
 * ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
 ENDERS · Applied Econometric Time Series
 † ETHIER and KURTZ · Markov Processes: Characterization and Convergence
 EVANS, HASTINGS, and PEACOCK · Statistical Distributions, *Third Edition*
 FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*, Revised; Volume II, *Second Edition*
 FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
 FITZMAURICE, LAIRD, and WARE · Applied Longitudinal Analysis
 * FLEISS · The Design and Analysis of Clinical Experiments
 FLEISS · Statistical Methods for Rates and Proportions, *Third Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- † FLEMING and HARRINGTON · Counting Processes and Survival Analysis
 FULLER · Introduction to Statistical Time Series, *Second Edition*
- † FULLER · Measurement Error Models
 GALLANT · Nonlinear Statistical Models
 GEISSER · Modes of Parametric Statistical Inference
 GELMAN and MENG · Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives
 GEWEKE · Contemporary Bayesian Econometrics and Statistics
 GHOSH, MUKHOPADHYAY, and SEN · Sequential Estimation
 GIESBRECHT and GUMPERTZ · Planning, Construction, and Statistical Analysis of Comparative Experiments
 GIFI · Nonlinear Multivariate Analysis
 GIVENS and HOETING · Computational Statistics
 GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
 GNANADESIKAN · Methods for Statistical Data Analysis of Multivariate Observations, *Second Edition*
 GOLDSTEIN and LEWIS · Assessment: Problems, Development, and Statistical Issues
 GREENWOOD and NIKULIN · A Guide to Chi-Squared Testing
 GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
- * HAHN and SHAPIRO · Statistical Models in Engineering
 HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners
 HALD · A History of Probability and Statistics and their Applications Before 1750
 HALD · A History of Mathematical Statistics from 1750 to 1930
- † HAMPEL · Robust Statistics: The Approach Based on Influence Functions
 HANNAN and DEISTLER · The Statistical Theory of Linear Systems
 HEIBERGER · Computation for the Analysis of Designed Experiments
 HEDAYAT and SINHA · Design and Inference in Finite Population Sampling
 HEDEKER and GIBBONS · Longitudinal Data Analysis
 HELLER · MACSYMA for Statisticians
 HINKELMANN and KEMPTHORNE · Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design, *Second Edition*
 HINKELMANN and KEMPTHORNE · Design and Analysis of Experiments, Volume 2: Advanced Experimental Design
 HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance
 of Variance
- * HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes
 * HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory Data Analysis
 HOCHBERG and TAMHANE · Multiple Comparison Procedures
 HOCKING · Methods and Applications of Linear Models: Regression and the Analysis of Variance, *Second Edition*
 HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
 HOGG and KLUGMAN · Loss Distributions
 HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
 HOSMER and LEMESHOW · Applied Logistic Regression, *Second Edition*
 HOSMER, LEMESHOW, and MAY · Applied Survival Analysis: Regression Modeling of Time-to-Event Data, *Second Edition*
- † HUBER · Robust Statistics
 HUBERTY · Applied Discriminant Analysis
 HUBERTY and OLEJNIK · Applied MANOVA and Discriminant Analysis, *Second Edition*
 HUNT and KENNEDY · Financial Derivatives in Theory and Practice, *Revised Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- HURD and MIAMEE · Periodically Correlated Random Sequences: Spectral Theory and Practice
- HUSKOVA, BERAN, and DUPAC · Collected Works of Jaroslav Hajek—
with Commentary
- HUZURBAZAR · Flowgraph Models for Multistate Time-to-Event Data
- IMAN and CONOVER · A Modern Approach to Statistics
- † JACKSON · A User's Guide to Principle Components
- JOHN · Statistical Methods in Engineering and Quality Assurance
- JOHNSON · Multivariate Statistical Simulation
- JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A
Volume in Honor of Samuel Kotz
- JOHNSON and BHATTACHARYYA · Statistics: Principles and Methods, *Fifth Edition*
- JOHNSON and KOTZ · Distributions in Statistics
- JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the
Seventeenth Century to the Present
- JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions,
Volume 1, *Second Edition*
- JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions,
Volume 2, *Second Edition*
- JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions
- JOHNSON, KEMP, and KOTZ · Univariate Discrete Distributions, *Third Edition*
- JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of
Econometrics, *Second Edition*
- JUREČKOVÁ and SEN · Robust Statistical Procedures: Asymptotics and Interrelations
- JUREK and MASON · Operator-Limit Distributions in Probability Theory
- KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
- KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence
- KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data, *Second
Edition*
- KARIYA and KURATA · Generalized Least Squares
- KASS and VOS · Geometrical Foundations of Asymptotic Inference
- † KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster
Analysis
- KEDDEM and FOKIANOS · Regression Models for Time Series Analysis
- KENDALL, BARDEN, CARNE, and LE · Shape and Shape Theory
- KHURI · Advanced Calculus with Applications in Statistics, *Second Edition*
- KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models
- KLEIBER and KOTZ · Statistical Size Distributions in Economics and Actuarial Sciences
- KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions,
Second Edition
- KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models:
From Data to Decisions, *Second Edition*
- KOTZ, BALAKRISHNAN, and JOHNSON · Continuous Multivariate Distributions,
Volume 1, *Second Edition*
- KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of
Time-Dependent Systems with Practical Applications
- KOWALSKI and TU · Modern Applied U-Statistics
- KVAM and VIDAKOVIC · Nonparametric Statistics with Applications to Science
and Engineering
- LACHIN · Biostatistical Methods: The Assessment of Relative Risks
- LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and
Historical Introduction
- LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE ·
Case Studies in Biometry

LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*

LAWLESS · Statistical Models and Methods for Lifetime Data, *Second Edition*

LAWSON · Statistical Methods in Spatial Epidemiology

LE · Applied Categorical Data Analysis

LE · Applied Survival Analysis

LEE and WANG · Statistical Methods for Survival Data Analysis, *Third Edition*

LePAGE and BILLARD · Exploring the Limits of Bootstrap

LEYLAND and GOLDSTEIN (editors) · Multilevel Modelling of Health Statistics

LIAO · Statistical Group Comparison

LINDVALL · Lectures on the Coupling Method

LIN · Introductory Stochastic Analysis for Finance and Insurance

LINHART and ZUCCHINI · Model Selection

LITTLE and RUBIN · Statistical Analysis with Missing Data, *Second Edition*

LLOYD · The Statistical Analysis of Categorical Data

LOWEN and TEICH · Fractal-Based Point Processes

MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in
Statistics and Econometrics, *Revised Edition*

MALLER and ZHOU · Survival Analysis with Long Term Survivors

MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel

MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of
Reliability and Life Data

MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets

MARCHETTE · Random Graphs for Statistical Pattern Recognition

MARDIA and JUPP · Directional Statistics

MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with
Applications to Engineering and Science, *Second Edition*

McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models

McFADDEN · Management of Data in Clinical Trials, *Second Edition*

* McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition

McLACHLAN, DO, and AMBROISE · Analyzing Microarray Gene Expression Data

McLACHLAN and KRISHNAN · The EM Algorithm and Extensions, *Second Edition*

McLACHLAN and PEEL · Finite Mixture Models

McNEIL · Epidemiological Research Methods

MEEKER and ESCOBAR · Statistical Methods for Reliability Data

MEERSCHAERT and SCHEFFLER · Limit Distributions for Sums of Independent
Random Vectors: Heavy Tails in Theory and Practice

MICKEY, DUNN, and CLARK · Applied Statistics: Analysis of Variance and
Regression, *Third Edition*

* MILLER · Survival Analysis, *Second Edition*

MONTGOMERY, PECK, and VINING · Introduction to Linear Regression Analysis,
Fourth Edition

MORGENTHAUER and TUKEY · Configural Polysampling: A Route to Practical
Robustness

MUIRHEAD · Aspects of Multivariate Statistical Theory

MULLER and STOYAN · Comparison Methods for Stochastic Models and Risks

MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and
Nonlinear Optimization

MURTHY, XIE, and JIANG · Weibull Models

MYERS and MONTGOMERY · Response Surface Methodology: Process and Product
Optimization Using Designed Experiments, *Second Edition*

MYERS, MONTGOMERY, and VINING · Generalized Linear Models. With
Applications in Engineering and the Sciences

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- † NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
- † NELSON · Applied Life Data Analysis
- NEWMAN · Biostatistical Methods in Epidemiology
- OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
- OKABE, BOOTS, SUGIHARA, and CHIU · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams, *Second Edition*
- OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis
- PALTA · Quantitative Methods in Population Health: Extensions of Ordinary Regressions
- PANJER · Operational Risk: Modeling and Analytics
- PANKRATZ · Forecasting with Dynamic Regression Models
- PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
- * PARZEN · Modern Probability Theory and Its Applications
- PEÑA, TIAO, and TSAY · A Course in Time Series Analysis
- PIANTADOSI · Clinical Trials: A Methodologic Perspective
- PORT · Theoretical Probability for Applications
- POURAHMADI · Foundations of Time Series Analysis and Prediction Theory
- POWELL · Approximate Dynamic Programming: Solving the Curses of Dimensionality
- PRESS · Bayesian Statistics: Principles, Models, and Applications
- PRESS · Subjective and Objective Bayesian Statistics, *Second Edition*
- PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
- PUKELSHEIM · Optimal Experimental Design
- PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
- † PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
- QIU · Image Processing and Jump Regression Analysis
- * RAO · Linear Statistical Inference and Its Applications, *Second Edition*
- RAUSAND and HØYLAND · System Reliability Theory: Models, Statistical Methods, and Applications, *Second Edition*
- RENCHER · Linear Models in Statistics
- RENCHER · Methods of Multivariate Analysis, *Second Edition*
- RENCHER · Multivariate Statistical Inference with Applications
- * RIPLEY · Spatial Statistics
- * RIPLEY · Stochastic Simulation
- ROBINSON · Practical Strategies for Experimenting
- ROHATGI and SALEH · An Introduction to Probability and Statistics, *Second Edition*
- ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance
- ROSENBERGER and LACHIN · Randomization in Clinical Trials: Theory and Practice
- ROSS · Introduction to Probability and Statistics for Engineers and Scientists
- ROSSI, ALLENBY, and McCULLOCH · Bayesian Statistics and Marketing
- † ROUSSEEuw and LEROY · Robust Regression and Outlier Detection
- * RUBIN · Multiple Imputation for Nonresponse in Surveys
- RUBINSTEIN and KROESE · Simulation and the Monte Carlo Method, *Second Edition*
- RUBINSTEIN and MELAMED · Modern Simulation and Modeling
- RYAN · Modern Engineering Statistics
- RYAN · Modern Experimental Design
- RYAN · Modern Regression Methods
- RYAN · Statistical Methods for Quality Improvement, *Second Edition*
- SALEH · Theory of Preliminary Test and Stein-Type Estimation with Applications
- * SCHEFFE · The Analysis of Variance
- SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
- SCHOTT · Matrix Analysis for Statistics, *Second Edition*
- SCHOUTENS · Levy Processes in Finance: Pricing Financial Derivatives

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- SCHUSS · Theory and Applications of Stochastic Differential Equations
- † SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
- † SEARLE · Linear Models for Unbalanced Data
- † SEARLE · Matrix Algebra Useful for Statistics
- † SEARLE, CASELLA, and McCULLOCH · Variance Components
- SEARLE and WILLETT · Matrix Algebra for Applied Economics
- SEBER · A Matrix Handbook For Statisticians
- † SEBER · Multivariate Observations
- SEBER and LEE · Linear Regression Analysis, *Second Edition*
- † SEBER and WILD · Nonlinear Regression
- SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
- * SERFLING · Approximation Theorems of Mathematical Statistics
- SHAFER and VOVK · Probability and Finance: It's Only a Game!
- SILVAPULLE and SEN · Constrained Statistical Inference: Inequality, Order, and Shape Restrictions
- SMALL and MCLEISH · Hilbert Space Methods in Probability and Statistical Inference
- SRIVASTAVA · Methods of Multivariate Statistics
- STAPLETON · Linear Statistical Models
- STAPLETON · Models for Probability and Statistical Inference: Theory and Applications
- STAUDTE and SHEATHER · Robust Estimation and Testing
- STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
- STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
- STREET and BURGESS · The Construction of Optimal Stated Choice Experiments: Theory and Methods
- STYAN · The Collected Papers of T. W. Anderson: 1943–1985
- SUTTON, ABRAMS, JONES, SHELDON, and SONG · Methods for Meta-Analysis in Medical Research
- TAKEZAWA · Introduction to Nonparametric Regression
- TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
- THOMPSON · Empirical Model Building
- THOMPSON · Sampling, *Second Edition*
- THOMPSON · Simulation: A Modeler's Approach
- THOMPSON and SEBER · Adaptive Sampling
- THOMPSON, WILLIAMS, and FINDLAY · Models for Investors in Real World Markets
- TIAO, BISGAARD, HILL, PEÑA, and STIGLER (editors) · Box on Quality and Discovery: with Design, Control, and Robustness
- TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
- TSAY · Analysis of Financial Time Series, *Second Edition*
- UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
- VAN BELLE · Statistical Rules of Thumb
- VAN BELLE, FISHER, HEAGERTY, and LUMLEY · Biostatistics: A Methodology for the Health Sciences, *Second Edition*
- VESTRUP · The Theory of Measures and Integration
- VIDAKOVIC · Statistical Modeling by Wavelets
- VINOD and REAGLE · Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments
- WALLER and GOTWAY · Applied Spatial Statistics for Public Health Data
- WEERAHANDI · Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models
- WEISBERG · Applied Linear Regression, *Third Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

- WELSH · Aspects of Statistical Inference
- WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for p -Value Adjustment
- WHITTAKER · Graphical Models in Applied Multivariate Statistics
- WINKER · Optimization Heuristics in Economics: Applications of Threshold Accepting
- WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
- WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
- WOODWORTH · Biostatistics: A Bayesian Introduction
- WOOLSON and CLARKE · Statistical Methods for the Analysis of Biomedical Data, *Second Edition*
- WU and HAMADA · Experiments: Planning, Analysis, and Parameter Design Optimization
- WU and ZHANG · Nonparametric Regression Methods for Longitudinal Data Analysis
- YANG · The Construction Theory of Denumerable Markov Processes
- YOUNG, VALERO-MORA, and FRIENDLY · Visual Statistics: Seeing Data with Dynamic Interactive Graphics
- ZELTERMAN · Discrete Distributions—Applications in the Health Sciences
- * ZELLNER · An Introduction to Bayesian Inference in Econometrics
- ZHOU, OBUCHOWSKI, and McCLISH · Statistical Methods in Diagnostic Medicine

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.